UNIVERSITY of OTAGO
TE WHARE WĀNANGA O OTĀGO

DUNEDIN NEW ZEALAND

Investigating Complexities Through Computational Techniques

Alec Holt

The Information Science Discussion Paper Series

Number 2000/01
February 2000
ISSN 1172-6024

University of Otago

Department of Information Science

The Department of Information Science is one of six departments that make up the Division of Commerce at the University of Otago. The department offers courses of study leading to a major in Information Science within the BCom, BA and BSc degrees. In addition to undergraduate teaching, the department is also strongly involved in postgraduate research programmes leading to MCom, MA, MSc and PhD degrees. Research projects in spatial information processing, connectionist-based information systems, software engineering and software development, information engineering and database, software metrics, distributed information systems, multimedia information systems and information systems security are particularly well supported.

The views expressed in this paper are not necessarily those of the department as a whole. The accuracy of the information presented in this paper is the sole responsibility of the authors.

Copyright

Copyright remains with the authors. Permission to copy for research or teaching purposes is granted on the condition that the authors and the Series are given due acknowledgment. Reproduction in any form for purposes other than research or teaching is forbidden unless prior written permission has been obtained from the authors.

Correspondence

This paper represents work to date and may not necessarily form the basis for the authors' final conclusions relating to this topic. It is likely, however, that the paper will appear in some form in a journal or in conference proceedings in the near future. The authors would be pleased to receive correspondence in connection with any of the issues raised in this paper, or for subsequent publication details. Please write directly to the authors at the address provided below. (Details of final journal/conference publication venues for these papers are also provided on the Department's publications web pages: http://www.otago.ac.nz/informationscience/pubs/publications.html). Any other correspondence concerning the Series should be sent to the DPS Coordinator.

Department of Information Science
University of Otago
P O Box 56
Dunedin
NEW ZEALAND

Fax: +64 3 479 8311
email: dps@infoscience.otago.ac.nz
www: http://www.otago.ac.nz/informationscience/

Investigating complexities through computational techniques

A. Holt
Department of Information Science, University of Otago
PO Box 56, Dunedin New Zealand
Email: aholt@infoscience.otago.ac.nz

Abstract
This article outlines similarity applied to the general environment and geographical information domains. The hypothesis is if physical and social sciences manifest similar amenities, then similarity would be a generative technique to analyse the cached information inherent in the data retrieved. Similarity is examined concerning the spatial grouping of natural kinds in a complex environment.

1 Introduction

Increasingly, scientific research is biasing towards computational approaches with the advances in technology. Supercomputing is now an accepted technique that stands as an equal partner to observation and experimentation. In practice, developing these new computationally dependent tools for analysis and modelling induce many new areas of computation. Environmental scientists are increasingly utilising information systems to help them make decisions in their very complex domain(s). Swayne suggests that a new applied discipline in Computer Science that started in the mid-1980's has since evolved dramatically into what he calls Environmental Informatics or Environmental Information Systems. An online air pollution monitoring system, for example, incorporates both informatics and environmental issues. Some fruitful informatics techniques that are being utilised for environmental management include; distributed database concepts, analysis techniques (artificial and computational intelligence), exploratory analysis (knowledge discovery and data mining) and visualisation techniques for modelling and simulation. GeoComputation, for example, is the word used to represent these changes of the past 5 years in the geographic information (GI) sciences (changes in the processing speed and the cost of high performance computing). The tools reflect the conglomerate technologies of high performance computing, artificial intelligence (AI) and GI science. If physical and social sciences manifest similar amenities, then this article suggests it would be a generative exploratory technique to analyse the cached information inherent in the similarity of the data retrieved. Especially taking note of the spatial grouping of natural kinds in a complex environment. This article outlines similarity applied to the general environment and geographical information domains.

> 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 understand objects, structures and actions existent in reality. Interpreting objects is a fundamental process of most human pursuits, and the idea that people classify together those things that they identify as being similar is both intuitive and utilised across a wide range of disciplines. Scientific theories can not be reduced to logical constructions of simple statements about the similarity between sense data, as many logical positivists thought, not even the 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 contribution of hypothesis in a new 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.

This article aims to emphasise that similarity could be used to search and solve patterns in the complexities of the natural world. This approach of modelling aims to analyse the environment in a holistic view rather than using the reductionist view (of using small components to explain a large ecosystem). In effect, by understanding the similarities (spatial or otherwise) between subjects, people can learn more about each subject than if we only studied the subject in isolation. Also this article aims to indicate the use of similarity in GI systems as the basis of a retrieval technique. A retrieval technique that can be used to sort things into groups whilst operating in an information science domain. The grouping and matching, for example, of topo-climates or indigenous flora and fauna in to their positions in the landscape.

2 Similarity

Our concept of similarity is 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 generalizations".

Quine construes a primitive form of comparative similarity that is an evolutionary product of natural selection. With respect to how similarity fits in with regularities of nature, to afford us reasonable success in our primitive induction's and expectation. 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 block 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 becomes multiple as one develops and matures, making perhaps for increasingly dependable prediction. Interestingly, our senses compliment new and novel grouping of kinds, they are not superseded. Quine suggests that, our 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. 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 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 (Quine 1969:136).

However, currently reductionism has fallen, and Green (1997) suggests complexity holds some answers to the functioning of the environment. This article proposes that the "drive for a cosmic similarity concept" can also identify with the complexity paradigm.

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 that 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. 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 used commonalties to indicate similarity, in saying that, "something runs through the whole thread - namely the continuous overlapping of those fibres" (Wittgenstein 1958:31-32). 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 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 (Popper, 1972:422).

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 this is an argument for his idea that 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, for example, of white swans is what makes people believe in the general theory that all swans are white. According to Popper, "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). Point of view in information systems can be construed as context. Context has a major influence on the type of information that can be retrieved using the similarity concept. With context, scale, 'techniques for retrieval' and 'similarity measures' have important roles in similarity.

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 a (for the time being) 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.

The grouping of animals in to their natural kinds is the example shown in the next section of this article. Where a set of animals in a zoo are used to test if CBR classify them correctly? The set of animals include 101 animals and range from an aardvark to a worm. Features such as hair, feathers, eggs, backbone, breathes and fins are used to classify the animals. A Wren, for example, is identified as a bird and is considered more similar to a Lark than to a Flamingo.

3 Mechanics of the similarity matching process

The similarity matching process in this article incorporates CBR techniques. CBR is a general paradigm for reasoning from experience. It assumes a memory model for representing, indexing, organising past cases and a process model for retrieving and modifying old cases and assimilating new ones (Klein *et al.* 1988; Schank 1982). More specifically, CBR is defined as:

> adapting old solutions to meet new demands, using old cases to explain new situations, using old cases to critique new solutions, or reasoning from precedents to interpret a new situation or create an equitable solution to a new problem. (Kolodner 1993:4).

3.1 Cases

Cases are the fundamental units of CBR. They are the essence of CBR and their structure in effect determines how CBR operates. Cases are unique chunks of information. A case is defined as "a contextualised piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of a reasoner." (Kolodner 1993:13).

3.2 The CBR cycle

Aamodt & Plaza (1994) suggest that a CBR cycle may be described by the following four processes:

1. retrieve the most similar case(s),
2. reuse the information and knowledge in that case to solve the problem,
3. revise the proposed solution,
4. retain the parts of this solved problem (experience) which are likely to be useful for future problem solving.

3.3 The similarity matching process

The case definition defines the field types, the formalism for the input values and the weights of the fields. The information in the case definition is used for type checking for the input cases, while the weights are used to aid the case-matching process. An example of the formalism of the case definition for the AIAI case file is shown in Figure 1. Indicating the importance of the field name *breathes with lungs* which has a *weight of 4* in comparison to the field *has fins* which has a *weight of 1*.

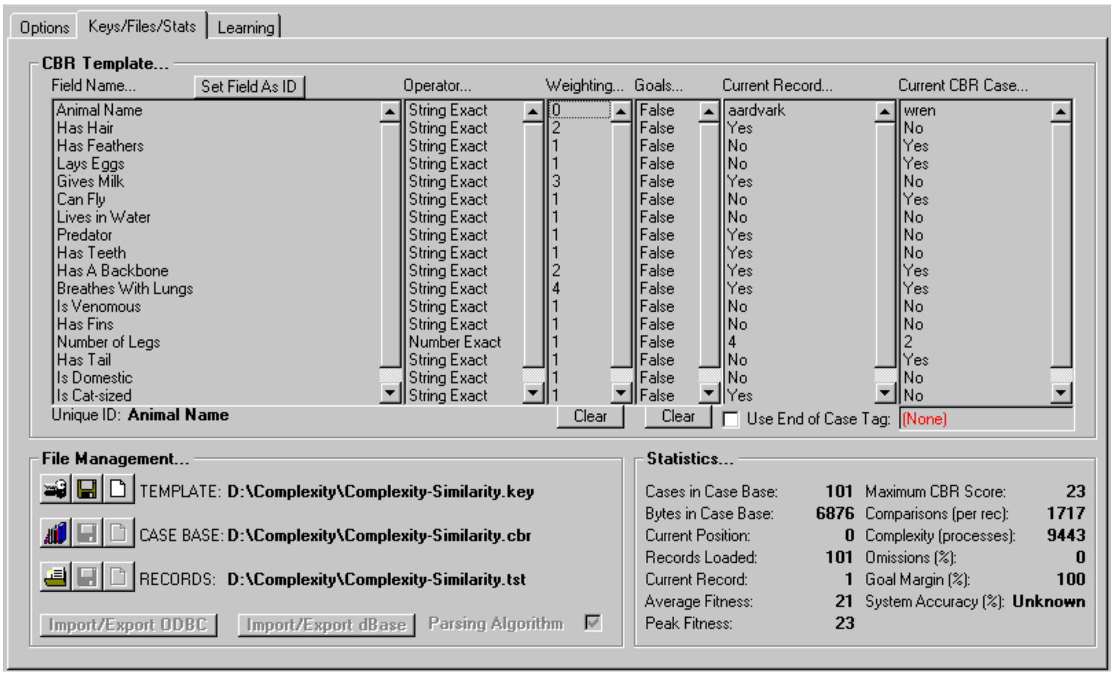

Figure 1. An example of the formalism of the case definition for AIAI

The index definition defines the fields used as goals (indexes) when searching for a matching case. A case base must have at least one field used as a goal.

Similarity matching follows the process below:

1) The user enters the criteria (CBR Template) to build the case structure. This is used by the case matching process, using goals and weights, to find a previous, similar case.

2) The system then performs a search (based on criteria provided by the user) and may find a subset of cases that match the index constraints. AIAI uses animal type as a goal, therefore grouping all cases with the same animal type. Then the system searches the subset of cases for the final case(s) that match all the criteria. The final case(s) are then matched and ranked.

3) If no case(s) match all the criteria (for instances when there are only a few cases in the case base) then the system prompts the user to change the goal in order to initiate another search. Goal constraints can be designed to be more general by specifying abstract values or fewer constraints this will increase the probability of the system finding matching cases.

4) The case instance with the highest weight score is selected as the highest ranked answer.
The following stages are used in the case matching and ranking process.

- Weight rules are triggered and added to the case match score.
- The goal values help retrieve a subset of cases from the case base which match all the goal values exactly (except when abstraction symbols are specified as goal values in the modification rules).
- Thresholds can be set (see Figure 2.).
- Different retrieval techniques are available.

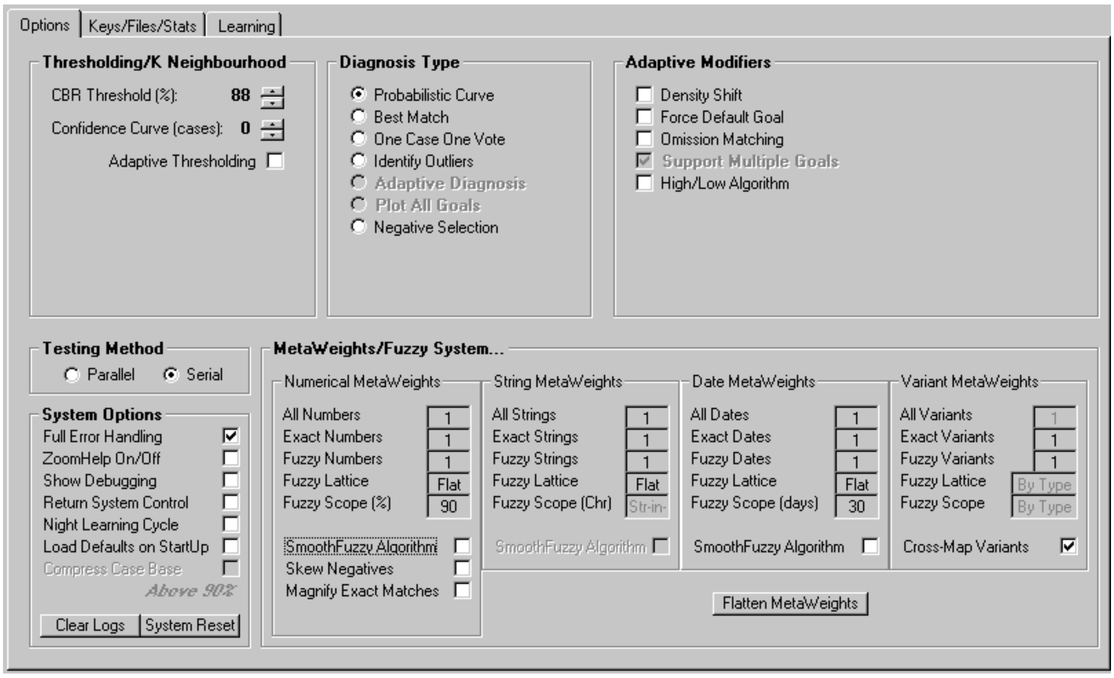

Figure 2. An example of the weights, goals and retrieval techniques available

- Once this list of cases has been retrieved the user can,
  (a) allow the program automatically (in batch-mode) to select a case based on weight-matching. For each case in the subset the case-matcher finds a weight which is obtained by totalling the weights of all the fields that matched. Fields which do not match exactly, but are defined to be similar by the modification rules, return a value which is less than the field's normal weight. The case matcher selects the case with the highest total weight. Alternatively the user can,
  (b) browse through the selected cases and select a case manually (compare case by case basis).

Methods for case retrieval include nearest neighbour matching, induction, knowledge guided induction, structuring using the interquartile distance, the k-d tree, similarity measuring in the k-d tree, exemplary 2-d tree, fuzzy algorithms and template retrieval. These methods can be used alone or combined into hybrid retrieval strategies. The retrieval method used in this article is the nearest neighbour matching technique. Nearest neighbour algorithms are executed in a common fashion and is represented in equation 1. (Watson 1997)

$$Similarity(T,S) = \sum_{i=1}^{n} f(T_i, S_i) \times W_i$$

Where: T is the target case,
S is the source case,
n is the number of attributes in each case,
i is an individual attribute from 1 to n,
f is a similarity function for attribute i in cases T and S,
W is the importance weighting of attribute i.

Equation 1. A typical nearest neighbour algorithm (Watson 1997:28).

The nearest neighbour approach involves the assessment of similarity between stored cases and the new input case, based on matching and ranking each field and the respective weights. The user decides if certain features need weighting and if they do what must be the various ratios between the weights of various features. A limitation of this approach is that retrieval times increase with the number of cases. Therefore, this approach is more effective when the case base is relatively small.

5) Repairs are carried out on the selected case. On occasions, additional information can be requested by the system to clarify a complex situation.

6) If the user is dissatisfied with the previous matching case(s) further cases may be examined. This is continued until they are satisfied or until the user exhausts all matched cases (see Figure 3.).

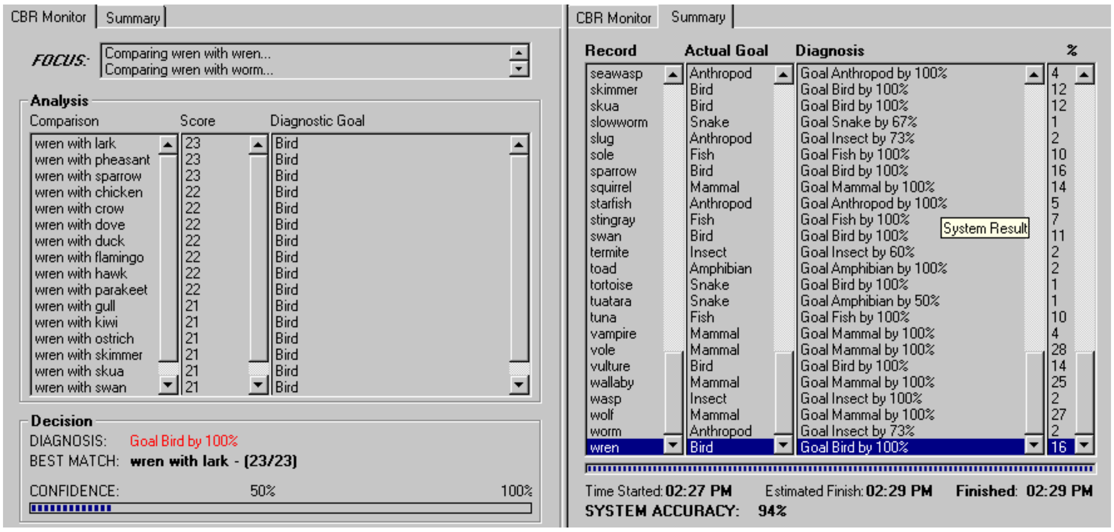

Figure 3. Examples of the similarity matching results

7) If the user is satisfied with the case selected and the solution offered then the new case is entered into the case-base (system learning).

With the advent of Environmental Informatics there are a variety of "things" that can be matched, for example, text, pixels, images, GI coverages, themes, layers, relationships, spatial attributes, processes. These can be matched for a measure of similarity at various dimensions and contexts. This article briefly outlines similarity for general environmental modelling and for GI science.

4 Allowing the environment to model itself

In trying to understand the complexities in the environment, every new piece of the puzzle discovered helps the subsequent pieces to be found and new domains to be explored, and confirms to or challenges existing theories. One complexity helps us understand another, for example, searching for volcanic action, and other geomorphologic earth traits on distant planets helps us understand other planets. This approach enables us to compare the unknown to what we already understand.

One of the important processes to the modelling of the environment is the way in which data is represented, structured and stored. One component of representing and structuring data is abstraction. Abstraction comes from the software-engineering school of thought to the modelling of a business (a closed, well known, created and simple system). Is it correct to use the same philosophy to an unknown and complex system like the biosphere or (which is more often the situation) a subset of the biosphere? Holt (1996) proposed an alternative approach by allowing environmental indicators to be used directly to model the environment. Rather than modelling the environment by abstracting, weighting and biasing information for the model, Holt (1996) allowed the environment to dictate its own software engineering, its own model. Holt used the environment as a search space for finding similarities. Within a complex community there are similarities, features are similar and features exhibit similar relationships. It is like hypothesising, have I seen this pattern before (using the brute force of a computer to search for a pattern), if so where and what was it, what was the context, what was the scale?

Bossomaier and Green (1998) suggest that the environment has a major influence on shaping natural communities. In particular they suggest that landscapes influence the ecology of an area. The simple examples they give are that temperature decreases the higher up a mountain we climb or rainfall runs off hilltops into gullies. Bossomaier and Green (1998) mention climate envelopes

> If we know the climate 'preferred' by a given species then we can predict its potential distribution by looking on a map for all sites in the landscape that fall within that envelope. By 'climatic envelope' we mean the range of values that the species tolerates for rainfall, temperature and so on. (Bossomaier and Green 1998:73)

It is suggested using *climate envelopes* various environmental habitats can be predicted. Holt and Benwell (1999) predicted the habitats of various soil types. Chunks (as a snapshot of the environment in a dynamic state) of the real-thing were used to formulate the cases in the case-base. Kolodner (1993) suggests CBR should be considered when it is difficult to formulate domain rules but cases are available. Holt and Benwells (1999) research involved CBR to retrieve, reuse, revise and retain previous similar cases. With CBR techniques soil scientists could refer to records of previous zones with similar input requirements and adjust the parameters for a similar zone to reflect the different requirements of the new zone being classified. The records of previous zones are good examples of working compromises between the different operating requirements. Attributes of the following features were measured, land element, slope, aspect, A_horizon texture, B_horizon texture, soil depth, A_thickness, dung, % gravel volume, % gravel weight, % carbon weight and % carbon volume. These features are important to soil forming processes and are controlled, to some degree, by landforms (Hewitt 1995). Consequently, it is suggested that soil classes should nest within specific landforms (due to the influence of the environment and landscape on the ecology). Holt and Benwell (1999) used the attributes of the features as fields and goals to define the case instances. These case instances store the fields that are used to match unclassified soil sites with previously similar fields which have a known classification. Holt and Benwell (1999) used this technique with a 88% accuracy success rate to classify soil types nested in various landscapes. The main characteristics of the soil classification problem are a large volume of observed historical data, the complex nature of the environment, the soil classification modelling rules for the decision making procedure, and the use of analogical reasoning by experts during problem solving. CBR offered advantages for all the features above: the historical data represents a repository of past experience that can be transformed into cases, and the abstract rules for modelling soil classification formed the structure of the cases by defining their structure and which fields could be used. More advantages of using CBR included the explanation and learning abilities provided by the system. Explanation facilities are simple to implement in a CBR system. These explanation features make the users more comfortable with reasoning by analogy because the justifications utilise data observations of past soil series incidents to support proposals instead of chain rules that are triggered by abstracted threshold values. This explanation gives a system more chance of being accepted, where ultimate responsibility for the decisions remains with the users. CBR systems can continuously incorporate new data as cases and, in this way, adapt to long-term trends such as soil degradation/soil loss or to other load systems.

Characteristics that make CBR a good option to solve environmental problems include;
CBR enables the ability to use explicit experiences to aid in soil classification. These experiences (cases) will enable the CBR to provide solutions (classification) to similar cases. CBR differs from algorithms in the way that no one solution is offered (as with an algorithm) and the user can then choose a similar case from the solution set. Also, algorithms need all criteria to be fulfilled before it is run (likewise a formula needs all its parameters to produce a result), in comparison CBR allows some fields to be left blank, without jeopardising the result. Another advantage of CBR is that its results are improved after each iteration of a new case being added to the case-base (as experience improves so will the ability of the system to provide the best solution), in comparison an algorithm will predict the same answers and the associated error level each time it is used. Environmental problems are inherently complex and Holt (1996) proposed a novel method (a GI-CBR hybrid), that will aid in the modelling and solving of such problems. Taking a snapshot and instance of a complex system to model similar complex systems. This is modelling approach not intended to be the complete answer but it does further our understanding of managing and modelling the complexities of the environment.

Holt and Benwell (1999) define spatial similarity as those regions which, at a particular granularity (scale) and context (thematic properties) are considered similar. Spatial similarity 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.

Holt & Benwell (1999) proposed a spatial similarity system 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. 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 spatial similarity system 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 is 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. Consequently, based on a set of criteria selected by the user similar instances can be found. It is not just the attributes that determine similarity, spatial relationships between situations also affect similarity. 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 GI system is fortuitous. A GI system 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).

5 Discussion

The author considers similarity assessment generative for retrieving and analysing complex environmental and spatial information. It may help researchers describe and explore certain phenomena, its immediate environment and its relationships to other phenomena. This article identified that phenomena are similar to each other depending on the type and number of commonalties they share and that the spatial properties of entities have an impact on similarity. The types of applications of similarity for the general environment and spatial domains are indicated in table 1 below. Similarity queries may be utilised in future applications to answer questions such as;

- Are there phenomena similar to the searched example?
- Which spatial phenomena have the following properties?

Environmental applications	**Spatial applications**
Matching the known geological features on earth with those observed on distant planets, and also using observations on distant planets for the better understanding of similar happenings on earth.	We can forecast the landscape of the future and re-create the landscape of the past through pockets of vegetation by modelling similar areas with similar landscapes.
Genetically similarity, traits, may help in the isolation of various diseases	Similar farms, in products and size can be matched to allow management procedures to be shared.
Areas of similar community relationships and ecological character	From digital terrain models, points, lines, polygons and pixels can be clustered into groups with similar spatial kinds and then displayed with GI systems.
Matching chemical and molecular structures	The spatial organisations and groupings at a molecular and chromosome level can be explored and understood based on similar shapes and structures.
Genetics and pharmacology relationships	Weather patterns can be predicted based on similar historical meteorological data in similar landscapes.
Matching chromosomes , the human genome and human behavioural and health traits	River coverages and flooded areas can be used to locate similar areas prone to flooding.
Matching the delivery action and the position on the pitch with the bounce of a cricket ball	

Table 1. Types of applications of similarity for the general environment and spatial domains

Note: Conversely dissimilar, divergent areas or patterns of outlying areas can be explored in the equivalent manner as searching for similar landscapes.

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

6 Acknowledgements

Richard Wheeler from University of Edinburgh for the use of AIAI software. Richard Forsyth the creator and maintainer of the Zoo database. Olle Järnefors for his comments.

7 References

1 Aamodt, A., and Plaza, E. (1994) Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *Artificial Intelligence Communications,* **7** (1) pp. 39-59.

2 Bossomaier T. and Green D.G. (1998) *Patterns in the Sand: Computers, Complexity and Life*. Allen and Unwin, St Leonards, NSW, Australia.

3 Couclelis, H. (1998) Geocomputation in context. *Geocomputation: A Primer*. (Eds.) Longley, P. A., Brooks, S. M., McDonnell, R., & Macmillan, B., (Chichester: John Wiley & Sons), pp 17-29.

4 Green D. G. (1997) Complexity in ecological systems. Chapter 18 of *Frontiers in Ecology: Building the Links*, The 1997 Conference of the Ecological Society of Australia, Charles Sturt University, Albury, 12 pages http://life.csu.edu.au/esa/esa97/papers/green/green.htm

5 Green D. G. and N. I. Klomp (1998) Environmental informatics - a new paradigm for coping with complexity in nature. *Complexity International* Vol. 6 http://life.csu.edu.au/complex/ci/vol6/green/

6 Hewitt, A. E. (1995) Soils of the Conroy Land System, Central Otago. *Manaaki Whenua Press, Landcare Research Science Series*, Lincoln, Canterbury, New Zealand.

7 Holt. A. (1999) Spatial similarity and GIS: the grouping of spatial kinds. *The 11th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, Dunedin, New Zealand, pp 241-250.

8 Holt, A. (1998a) An environmental perspective on the affects of rapid changes in information and computation techniques. *Environmental Perspectives*, Issue 20:8-9.

9 Holt, A. (1998b) The Matching and Ranking of Surface and Deep Similarities of Spatial Data. *The 10th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, Dunedin. pp. 133-143.

10 Holt, A. (1997) GeoComputation - neologism, or gambit towards progressive research *Environmental Perspectives*, Issue 15:5-6.

11 Holt, A. (1996) Allowing the Environment to Model Itself. *Environmental Perspectives*, Issue 10: 6-7.

12 Holt, A. and G. L. Benwell (1999) Applying Case-based Reasoning in GIS. *The International Journal of Geographical Information Science*. 13(1):9-25.

13 Klein, G. A., Whitaker, L. A., and King, J. A. (1988) Using Analogues to Predict and Plan. *Proceedings of the Workshop on Case-Based Reasoning,* Clearwater Beach, Florida, ed. Kolodner. J., San Mateo, California. pp. 224-232.

14 Kolodner, J. (1993) *Case-Based Reasoning*. Morgan Kaufmann Publishers, San Mateo, 668 p

15 MacGillavrey C., J. B. Wilson & A. Holt (1999) The Use of Artificial Intelligence techniques and GIS for predictive vegetation modelling. *The 11th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, Dunedin, New Zealand, pp 305-306.

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

17 Popper K. R. (1972) *The Logic of Scientific Discovery*. London: Hutchinson 480 p

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

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

20 Schank, R. (1982) *Dynamic Memory: A Theory of Learning in Computers and People*. New York, Cambridge University Press.

21 Swayne, D. (1998) Towards a Computing Curriculum for Environmental Science. *Computer and Information Science Seminar, (William Evans Visiting Fellow)*.

22 Watson, I. D. (1997) *Applying Case-Based Reasoning; Techniques For Enterprise Systems*. San Francisco, California: Morgan Kaufmann Publishers.