Emergent Relations

Self-Indexing Media for Case-Based Reasoning

Milton TAN

Ministry of Information, Communications and the Arts, Singapore, and Department of Architecture, National University of Singapore, Singapore

Keywords: case-based reasoning, creativity, emergence

Abstract: A priori indexing systems for case-based reasoning are unlikely to engender interesting, let alone creative, explorations. The problem is that they assume a predetermined knowledge and conceptual structure which neglect new and emergent relations. By allowing multiple labelling of entities, it can be shown that different orders of emergent relations can be derived on the fly without any conventional classification schemes. This gives a flexible, scalable, and most importantly, an associative reasoning system that is more likely to trigger interesting concepts which are technically "categorical errors" but seeds of creative thought.

1 INTRODUCTION

This paper introduces a novel self-indexing system based on ideas and theories of natural emergent behaviours found in the connected lives of ants, brains, cities, and software (Johnson 2001). It is an elaboration of an idea first discovered by the author in 2000 (Tan, Tan and Ngahtemin 2000) in attempting to develop a search engine that moves laterally rather than vertically up and down fixed classification hierarchies.

Such a recombinant emergent system for creatively spinning knowledge webs is not known to exist. This paper draws generally on key issues of emergence and focuses on the structure and operations of a prototype self-indexing system. It is therefore not an exhaustive critique of emergence or case-based reasoning (best left to dissertations and text books).

One of the difficulties in maintaining a case-based repository is indexing. Without a good indexing system, the collection is as good as a freeform lifestyle magazine for casual browsing. The accepted technique to deal with this is the veritable library classification system. It is an autocratic method as an indexing system, as well as in implementation — i.e. the method is established ahead in time irrespective of the kinds of material that may turn up for classification. With experience and feedback, the methods gets to be increasingly more accurate in 'anticipating' the labels that would be required for putting away the media in such a way that it would be easily retrieved in future. Hopefully.

The conventional top-down indexing system assumes the following: a) Indexing systems assume completeness, needing only better resolution, b) Nothing is so new that it will require new categories, c) Material in the repository do not ever need to be re-classified. In sum, conventional indexing systems assume a stable world where categories and concepts to organise the objects and entities in the world are absolutes. This is perfectly acceptable until one begins to "think outside the box" in different "worlds" (Goodman 1997) — an implicit objective in case-based reasoning — to break new grounds and make paradigm shifts.

Ants, brain cells (and any organic cell for that matter), day-to-day entities in a city, and much of software systems such as the World Wide Web, are not centrally or top-down controlled, as we may assume. The so-called "Queen ant", unlike most queens, does not rule over her colony; she is more akin to a factory to increase its population. Our visual perceptual system is not controlled by some equivalent of a TV producer. The resulting systems are therefore determined by bottom-up codes that ordinary members of the system understand and apply at local points-of-contact without recourse to "master plan" or pace setters.

The prototype self-indexing system, takes after the idea that an effective and flexible system can exist, grow, and adapt from basic labelling embedded in the low-level discrete entities of the repository; with no need for a top-level pre-determined system. The system employs a simple method of labelling or "tagging" media files by recording keywords or symbols (after all, words are also part of a symbol system) for each discrete entity. The list is unlimited, and can be modified at any time. Multiple lists can also co-exist.

Since there is no overarching system to determine the relations of tags, the effective index system will have to grow out of the labels embedded in the collection. Firstly, the multiple labels of a particular entity of the repository collectively represents, perhaps for the first or only time, the relationship of these labels in that particular combination. Any subsequent match to this combination would be deemed as identical members of the same category. Schemes of partial or weighted match can be made to locate more 'distant cousins'. Secondly, it can be shown that by partial matching between entities, a trace network of the emergent index system can be derived. This represents the indirect relations of labels due to some common relations between entities.

2 INDEXING PROBLEMS IN CASE-BASED REASONING (CBR) SYSTEMS

"Scotch", "Cold Hands" and "Bad Knee" are nicknames of a group of friends who only meet once a year in a village in Switzerland to ski. The nicknames help them remember each other by their respective unique stories. Scotch once took off in a hang-glider held together with Scotch tape meant to have been replaced with proper fastenings; Cold Hands leaned on a boiler-stove in conversation and melted his prosthesis hand; and Bad Knee earned a questionable reputation as a regular skier on medical leave for a knee injury.

Names differentiate. Every time we give something a name we set it apart from practically everything else. Naming is also performed casually and even involuntarily. Our perceptual system has no way of determining absolutely everything about everything before we can reason about the world around us and decide what to do next. "Our vision is constructed" (Hoffman 1998). Harvard Philosopher Hillary Putnam grappled with the scientific "R" and common sense "r" of reality in his "Many Faces of Realism" (Putnam 1988a), and "Representation and Reality" (Putnam 1988b) and argued that the way we relate things in common is far from scientific. E.g., things that are red in common (red apples, red star, red light) defy scientific grouping. If, therefore, an external and absolute system of indexing the world to establish "things in common" is problematic, a different approach will be needed. For this, three key assumptions in conventional indexing systems need to be overturned:

Indexing systems are complete, needing only better resolution

Nothing is so new that it will require new categories

Materials in the repository do not ever need to be re-classified

The new approach aims to establish a self-indexing system that does not depend on a priory classification schemes (such as a library system), but one which would accommodate new and implied categories "on the fly", and enable old entities to be regrouped. Unless these new functions are developed, we would continue to be straight jacketed within tailor-made worlds, and with limited scope for thinking differently and hard pressed to find new and creative connections because they would technically be "categorical errors".

3 MULTIPLE LABELS

Assume for now that labels can be attached to CBR entities, not unlike a luggage tag for a suitcase. (What comprise entities and labels in practice, and how is it applied and used will be dealt with later.) Assume further that an unlimited number of unordered labels can be used. This is contrary to a library system where a book (the entity) will be given a classification code (the label) for one 'correct' location only. However, the non-library method is consistent with the way we ordinarily ascribe attributes and nicknames over and above a person's formal name. We can, abstractly, represent an entity as a closed shape, such as a circle or polygon; and its labels as marks or symbols within its boundary (Figure 1). For now, the shape or size of the entity does not matter. The distinguishing marks or symbols for labels are associated with the entity they refer to. An entity comprises a set of labels.

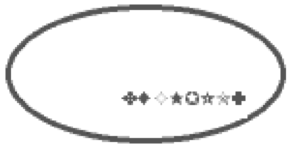

Figure 1 A Representation of an Entity

An architectural example of this is the multiple labelling of the image of the "Bat Chum" shrine (Figure 2), where the 15 labels — from "articulated pilaster" to "tantraka" — were applied to the entity by a domain expert. These labels are the equivalent of the symbols in Figure 1.

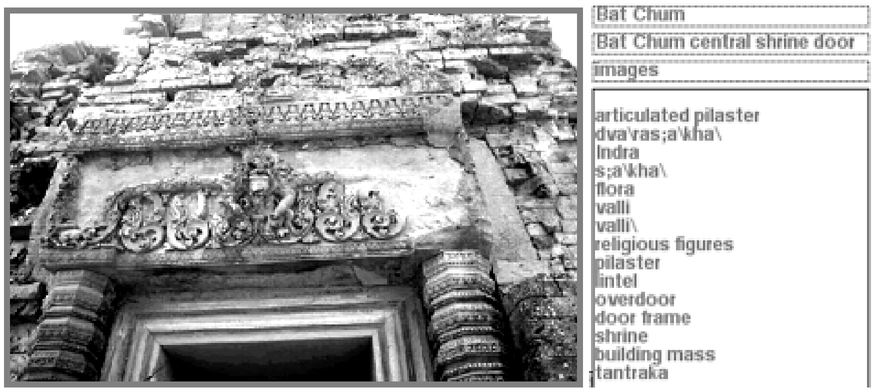

Figure 2 The Multi-labelling of a "Bat Chum" Entity (Indorf 2000)

To be instantiated, an entity must have at least one label. Different entities can, of course, contain different sets of labels, some or all of which may be common (Figures 3 and 4).

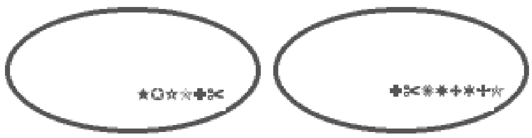

Figure 3 Two Different Entities with some Identical Labels

An entity can change its set of labels to respond to circumstances by the addition or subtraction of labels, or even by a complete substitution. The entity that we label "Car" or "Automobile" today was once labelled "Horseless Carriage". The "Radio" was once the "Wireless Set" and even the "Transistor".

The simple provision of multiple labels opens up the world of emergent relations in two ways:

Intra-entity: the individual labels of an entity relate to each other by being first order members of the same family of labels

Inter-entity: the set differences of two or more entities with at least one common label form the second order members of the family. From Figure 3, this is therefore ★✪✫✫❋✱✚✱✚✫ (✚✂ being the intersection). The intersection of the "Bat Chum" and "Preah Ko" entities (Figures 2 and 4) is "pilaster, overdoor, door frame, shrine, building mass".

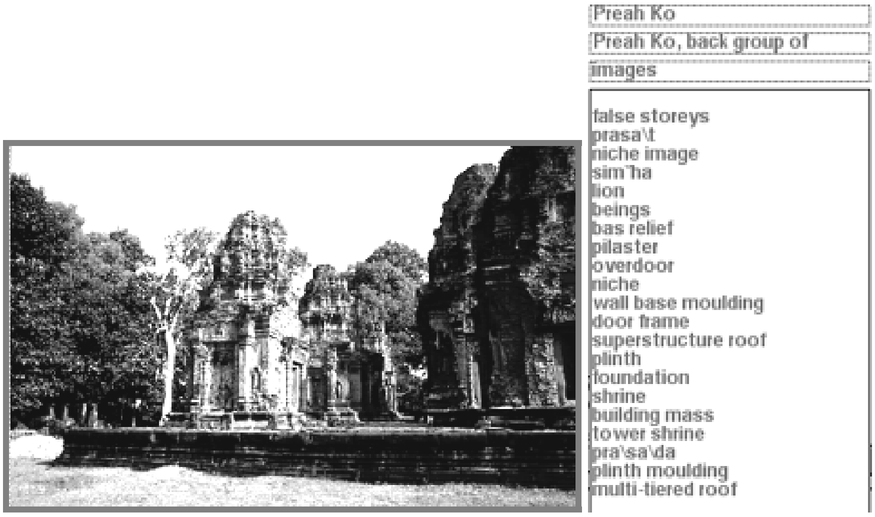

Figure 4 The Multi-labelling of a "Preah Ko" Entity (Indorf 2000)

We shall now see how the emergence works.

4 CLASSIFICATION SYSTEM ON THE FLY

Two types of emergent relations are produced from any collection of entities arising out of Intra-entity and Inter-entity operations. Intra-entity emergence produces "first order" relations whereas inter-entity emergence produces "second order" relations.

4.1 Intra-entity Emergence

The individual labels within an entity are related to each other on par. The entity legitimises the official first order relationship of these labels as equal members of a family. Even if the entity is a unique, one of a kind, instance of this particular set of labels, the association hold.

From the "Bat Chum" shrine in Figure 2, some of the first order emergent relations of the entity are:

“Indra” and “Valli”

“Shrine” and “Tantraka”

“religious figures” and “articulated pilaster”

“door frame” and “lintel”

With the instantiation of a universe of just one entity with a minimum of two labels, a classification system immediately emerges which will maintain that any entity with one of the labels will automatically be associated to any entity with either of the labels.

For example, if an existing entity has “cart” and “horse”, a new entity with “horse” (and perhaps several other labels) will, as a result of the emergent association of “cart” and “horse”, be open to further conceptual connections via the adopted label “cart”.

4.2 Inter-entity Emergence

Arising out of the introduction of this second entity, the original and further entities with “horse” will also have associative pathways from all the other labels in the second entity. For example, if the second entity had “horse”, “saddle” and “stable”, the emergent classification would also suggest an association between “cart” and “stable”, and between “cart” and “saddle”.

If the “Prasat Kravan” entity (Figure 5) is introduced into the collection along with “Bat Chum”, a series of second order emergent relations is formed. These include, however unreasonable or apparently meaningless:

“false storeys” and “religious figures”

“flora” and “dva\ra”

“wall base moulding” and “s;a\kha\”

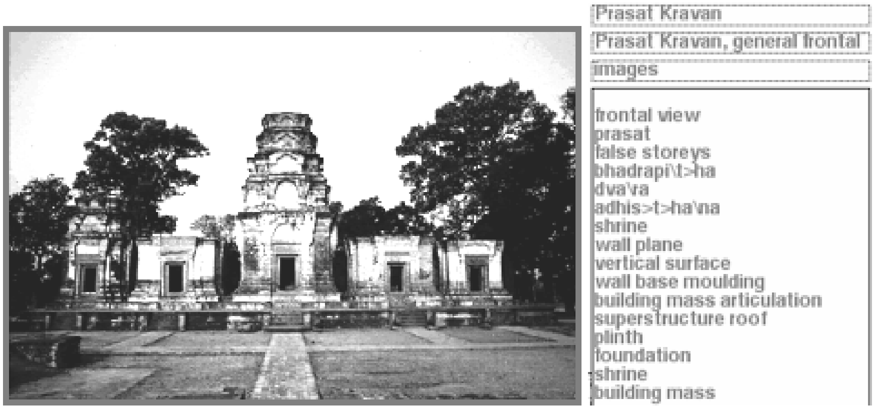

Figure 5 The Multi-tagging of a "Prasat Kravan" Entity (Indorf 2000)

If a third entity has "cart" and "apples", the second order emergent relations between "horse" and "apples" will arise, along with "saddle" and "apples", and "stable" and "apples". Is this opening up an accuracy-of-search problem? Yes, if the expectation is convergent search for categorical fidelity (of the yellow pages type). No, if the plan is precisely to loosen up the associative pathways to allow for the less likely and tangential connections which may offer interesting possibilities. Inter-entity emergence offers relations between the predetermined classifications of a priory systems and the utter random chances from no system.

How loose can the search be?: As loose as the number of levels or orders of labels. Following the above example, if an additional entity contains "apples" and "pie", a third order association is "horse" and "pie"! Whilst this may sometime appear ridiculous, there is a method in the madness. In this case, the link is "horse" to "cart" to "apples" and to "pie". Such associative reasoning is not unlike the powerful episodic reasoning that we frequently use to solve complex and demanding cognitive tasks.

How tight can the search be? As tight as bank accounts, in that every label of an entity, and perhaps even its sequence, much match exactly; one for one, like a combination lock, with no consideration for pathways and secondary links.

5 KEY DEVELOPMENT ISSUES

Two key development issues have arisen out of the project. The first concerns the strengths of relations between labels to provide differentiated pathways during search. The other concerns the value and management of the dynamic behaviours of the emergent system.

5.1 Powerful Associations

The emergent relations that can be derived from the multiple labelling of entities does not stop with the pathways. A collection of entities will also self-generate important indications of associative strengths. This can be used to prioritise and weigh options, and help guide and focus attention.

Associative strengths comes from:

- The number of occurrences of pairs of related labels, adjusted for levels or orders of association
- User feedback, passively from choice and speed of search patterns, and actively from ratings of associations and training of the system
- Explicit weighting and associations by experts and other stakeholders, whose recommendations are available by subscription

Associative strengths are independent of the pathways, and are therefore scalable and subject to separate fields of research and development. The emergent relations system provides a useful and practical framework for such investigations.

5.2 On the Fly

One of the most beneficial attributes of an emergent system is dynamic adjustment.

Labels can be altered on the fly, temporarily or edited to reflect new perspectives or circumstances. Different customised schemes can co-exist because the entities and their label sets can be stored separately in an information system. A practical use of such a function is expert or teaching systems in which the user, like a tourist, expects knowledgeable guidance.

More importantly, emergent systems recognise that the conceptual world is absolute, and therefore inherently unstable.

Emergent systems are also inherently scalable. The classification systems grow with the collection and adjust to its resolution of details as labels are edited as the collection grows. Associative strengths of relations can also be tweaked on the fly to interactively sculpt the exploratory space.

6 CONCLUSION

A conventional top-down index system can in fact be generated bottom-up by a self-indexing system which relies — like viruses — on the low-level attributes of local entities. It uses the emergent relations found by combining multiple labels within the entity (first order) and between entities with bridging labels (second order). The emergent system exhibits the interesting ability to spawn new categories and generates surprising new relations that would support creative exploration of ideas emerging out of relations rather than the parts of entities. The practical application of this system is in case-based reasoning of multimedia collections, particularly for creative or lateral exploration of ideas.

The project has opened several academic and commercial opportunities. The immediate key development issues are in the associative weighting of search pathways and in the management of the dynamic nature of the emergent system. Other issues include the incorporation of other forms of labelling such as symbolic, colour, graphics and gestures; and the extension into interactive user generated graphics.

(The development of the emergent relations system is a spin off from the OASIS (Online Architectural Studio Information System) research project R-295-000-007-112 funded by the National University of Singapore. Because a commercial version of the system is being developed, some critical intellectual property aspects of the system cannot be released at this stage.)

REFERENCES

Goodman, N. 1976. *Languages of Art*. Indianopolis: Hackett.

Hoffman, D. 1998. *Visual Intelligence*. New York: Norton.

Indorf, P. 2000. Report on SEACOR. *South East Asian Corpus: Historical Shrines.* Singapore: CASA.

Johnson, S. 2001. *Emergence*. New York: Scribner.

Putnam, H. 1988a. *Many faces of Realism*. Cambridge MA: MIT press.

Putnam, H. 1988b. *Representation and Reality*. Cambridge MA: MIT press.

Tan, M., Tan, B.K, and Ngahtemin, J. 2000. By Rhyme or Reason. Rapid Design Thinking by Digital Cross Referencing. *Proceedings of CAADRIA 2000*, eds. B-K. Tan, M. Tan and Y-C Wong, 399-410. Singapore: CASA, National University of Singapore.