Reprint NO.60

CI/SB (Agm) (L31)

UDC 681.3:699.82

THE KNOWLEDGE BASE - WILL THE FROG TURN INTO A PRINCE?

H.A. Trethowen

Reprinted from the, *"PROCEEDINGS OF THE SECOND NEW ZEALAND CONFERENCE ON EXPERT SYSTEMS, NZES'87"*, P 103. Conference Centre University of Auckland (2-4 February 1987)

BUILDING RESEARCH ASSOCIATION OF NEW ZEALAND BRANZ

THE KNOWLEDGE BASE

- Will the frog turn into a prince:?

H.A. Trethowen

Building Research Association of New Zealand (BRANZ)

Abstract: Experience in developing the "DAMP" diagnostic knowledge base has exposed difficulties in the process of transforming existing knowledge into usuable rules. Whilst some success has been achieved, it has also become clear that the gap between existing, intuitive knowledge, and the required rule-based form, is very great. There must now be an iterative process of recrystallising this knowledge in its new form, and/or a quantum step in computer tools to handle uncertain data in an intuitive-like manner, before knowledge-based systems can realize their potential.

This paper endeavours to define and illustrate the nature and difficulties in the knowledge transformation process, encountered with "DAMP".

1. Introduction

The reasons for interest is "expert systems" is that they may be able to make experience available to the inexperienced, and that they will perform consistently without overlooking facts or interpreting them differently from one time to another. In the case of diagnostic systems, these advantages come at a cost of incomplete generality and reliability, because the knowledge is imperfectly developed.

This paper aims to outline the nature of the DAMP Knowledge Base, and to examine some of the difficulties in transforming previous experience into that knowledge base. Although the paper describes the authors own experience, similar issues have been encountered by associates in other technologies - paint failures, weather penetration, building sealants. The possibility of, and reasons for, distortion in that transformation should be of concern by anyone involved with expert systems. The test of success in an expert diagnostic system must be an overall one, based on how reliably a user with a set of symptoms will be led toward a correct conclusion before they retire frustrated, and not on narrower technology factors.

There is a trend now to refer to "knowledge-based" systems rather than to "expert" systems. This is a healthy step, but does not go far enough. "Knowledge" exists in many forms, and the form required for computer application at present is as sets of rules. This form can be conveniently described as a "rule-based" system. The term "knowledge engineering" as used currently seems to refer to the processes of how best to handle these rules, once they exist. The interest in the first part of this paper is in the step prior to that, of establishing valid rules from the previous experience available.

2. Human v Machine Knowledge

Before discussion the DAMP knowledge base, it is worth reviewing some of the major differences between the existing experience-based knowledge of human experts, and the rule-based forms required for machine implementation. A comparison of the relative strengths of these two forms is suggested in Table I. Important differences come under several major headings:

(a) Theoretical Base

In tackling each new case, a human expert has at call not only a personal history of experience and published experience, but also background theories, hypothesis, or models on the processes of interest. In DAMP a case involving mildew growth would be viewed knowing that mildew can only grow where there is sufficient moisture, and a particular pattern of mildew growth has to be consistent with identifiable local factors which could cause moisture to be sufficient for mildew to grow at some points and not at others only centimetres or metres away.

This support is not available to a typical rule-based system such as DAMP. Consequently DAMP cannot assess in a particular case whether there is any feature of the mildew pattern which may alter the normal conclusions which are set out in its rules. Such exceptions do occur from time to time.

(b) The Knowledge form

One of the major hurdles in establishing rule-based knowledge, is that the form of such knowledge is totally different to the experience knowledge it is to be formed from. This makes it difficult for human experts to communicate their knowledge to a rule-based system.

The solution of many diagnostic problems rests on some pattern-recognition process. The human expert utilises effortlessly a degree of pattern-recognition skills that present computer technology does not approach. Worse, the patterns that may need to be recognised are not always clear to the human expert - they just "know" when they "see". Sometimes the patterns are verbal, or visual, but even this is not always true, and people seem to use mental imagery processes which do not correspond at all to any sensory medium. This sort of intellectual knowledge is extraordinarily difficult for the possessor to even identify, even harder to transfer without distortion.

Consider this sort of problem in regard to a successful trout fisherman, who is also willing to divulge the secrets. You can follow the expert angler's rules, but there is no certainty that you will catch anything. Why aren't there more millionaires, when there are so many good "How to be a millionaire" books? One needs knowledge in order to be able to make use of knowledge.

(c) Uncertainty

Humans generally, including experts, are quite accustomed to receiving information ranging from accurate to misleading to totally false, and sifting this subconsciously to extract the true and the useful. The present era of expert programs are not equipped for this. They must in general treat answers as true. To some extent the knowledge base can be set up to allow for inaccurate answers by seeking some degree of redundancy, and/or by testing for inconsistency. However the scope for these measures is restricted.

The significance of various clues varies according to circumstances, and there is a need for recognising this. Weighting systems using some probability factor have been considered for DAMP, and rejected. One reason for rejection is that the significance of various factors is strongly non-linear, and probably not consistent. For instance, the possibility that stains on the ceiling may be a result of moisture condensing under the roof is slightly increased if the stains occur at specific locations, or if they increase during or after cold weather. But if these conditions both occur together then the probability switches to very high.

It is a fact of life that client answers are not totally reliable, even when there is no intention to mislead. There are always plenty of people who don't know, will guess, or just try to fool the computer. What is seen as a "regular pattern" by an expert may not be seen as such by a client, especially if parts of the pattern are missing or concealed. In a 1984 survey of thermal insulation performance BRANZ field staff encountered quite some ignorance of owners concerning their houses, even on some fact they had reason to know. One owner assured our field staff that his ceiling had pink batts - he had personally looked. But there was no insulation of any kind. And this was not an isolated case.

It must also be kept in mind that people may not even be trying to be honest. Not a few building moisture investigations have been found to point to a rent dispute with a landlord, or a wish to deflect a threatened law suit on to another party, and their responses may be manipulated in a direction supporting their particular interests.

It is too much to expect today's expert system to be worldly-wise in such issues, but they may need to become so if they are to compare with human experts.

d. Judgement

In deciding how important a particular fact may be, the human expert is again at an advantage. Many moisture problems in buildings are diagnosed -accurately - on the basis of a mere handful of facts. But that handful comes from a pool of thousands, and is a different handful for every case. People can make a selection like this quite effortlessly, without even realising that they have executed an extraordinary trick. The process is laborious for a rule-based system. The real problem is to then verify that the rules always produce the right answer.

The process is paralleled in the cartooning field. All the detail and the million-bit complexities in a face, can be represented by a cartoonist with just a few pen-strokes. But they're a different set of pen-strokes for each face, and for each expression.

e. Future Development

If the strengths of human experts are in pattern recognition and general knowledge, a rule-based system has its strength in its ability to provide an alternative to personal hard-won experience. The number of good human experts is very small, the quality of many is suspect. The pseudo-experience obtainable from rule systems is evidently inferior to the best human experts, but it may be much better than misguided experience from a mediocre human.

Another substantial strength of computer based systems is that they can keep a virtually infallible record of each consultation. This may be important both for developing the knowledge system, and for defence against litigation.

A third and crucial reason for persisting with rule-based knowledge forms is their probable role in better direction of future knowledge as it evolves. By providing a specific framework for the knowledge, new (and old) data should become better assessed. Whether new data reinforces the rules or requires a change to the rules, its implications will be sharpened. I suggest that this effect on people is more important than has been recognised.

3. The DAMP Knowledge Base

: "Water is neither created nor destroyed (- building moisture always comes from somewhere)"

: "Moisture is the necessary and sufficient condition for decay, corrosion" Anon.

Given the ubiquitous nature of moisture problems, and the observed fact that rather few people seem adequately to understand the behaviour of moisture, we may start questioning whether there is any prospect for establishing a rule-based diagnostic system.

There is some historical evidence that it should be possible. BRANZ 20y+ experience began with published literature, much of which is sound but some quite defective. A long-standing advisory service using human "experts" now handles over 500-1000 N.Z. enquiries annually on moisture problems. At this level of enquiry it becomes possible and necessary to look for common factors, and to exploit them if present. It has been noted that in at least some cases these people could produce apparently correct diagnosis from brief descriptions. However, it is also noted that such cases often invoked some call on an extensive knowledge of building practice and occupant habits, sometimes indicated by minor cues in the description. The latter is difficult to pin down, as the "expert" is usually unaware of having done any such thing.

"DAMP" has already evolved through several stages. In one form it was a student project on Expert Systems (Sachdeva 1985). At present a more comprehensive version is being developed at BRANZ, using the CLASS language (Dechapunya, 1986, M. Buis et al 1986).

Conversion of this associative or intuitive knowledge base into rules, began with a separation into 11 building types, 8 symptom types, 6 sources of water, and 6 moisture transfer mechanisms. A schedule of these groups is given in Table 2, and a diagnostic tree illustrating the general scheme is given in Fig. 1.

Rules have been established to date for only one building type - domestic. Simple combinatorial expansion shows that there are about 300 possible combinations. However, potential users are not expected to know much about moisture sources or transfer mechanisms. But they can be reasonably asked about the patterns applying to their symptoms (eg, where the symptoms are seen; if, how, when they vary) and some information about the building itself, and about the way it is used. The number of possible combinations using this data promptly runs into thousands.

DAMP is fortuitously saved from being overwhelmed by the fact that the matrix of combinations is only sparsely filled. Only some 10% - 25% of the possible combinations are considered to exist and to add usefully to diagnosis.

The DAMP Knowledge Base firstly divides the moisture problems into three groups -

- mildew &/or surface condensation
- stains
- rot

Guidance is also available to help a user decide which if any of these conditions they may have. (This illustrates the beginning of the uncertainty - if someone is not clear what the problem is, the chances of getting accurate symptom descriptions must be suspect also).

In the case of mildew/condensation, the age of the building is established, and then the location within the building, Age is important only to establish whether or not the building is "new". If so, the chances are that the problems arise from construction moisture, ie, the use of timber or concrete which was installed wet or has become wet from rain, dew, or spillages during construction. These reservoirs of moisture in most cases dry off in the first few months of use. Furthermore, it is not usual for other, occupant-generated, sources of moisture to have accumulated sufficiently to be troublesome at that stage. Division into various locations in the building follows. As one might expect, problems in one room will not usually have their origin in another, although it is possible for more than one room to be affected by some common factor, notably the degree of heating and ventilating. These two items are key factors, and are the hardest to establish adequately. Occupants will clearly have no knowledge or interest in the real causative issues of mildew/condensation, which relate to the temperature differences between indoors and outdoors, the amount of indoor moisture their actions release, the amount of ventilation and the insulation level. Instead, one must ask indirect questions concerning heating and ventilation practices, and make assumptions about how equipment is probably being used. DAMP has been designed to default to an assumption that heating and ventilation are not sufficient if there are not adequate alternate indicators.

Although it is usual for problems in one location not to have their origins far away, there are some notable exceptions. Leaks have always been notorious for appearing far from their origin, but at least they are at a lower height than the origin. But there is a class of problems leading to stains, mildew, or dripping on ceilings, where roof moisture comes from subfloor ground. This can occur in buildings with masonry veneer or stucco walls, or with through-vented linen cupboards, that allow subfloor air to pass to the roof space. In cold clear weather, this process carries huge quantities of water into the roof. DAMP is set up to detect these cases, giving appropriate repair advice.

The "stains" and "rot" options do not bother about building age and go straight to identifying the location of the problems. These options are not yet as fully developed as the "mildew" option. The "rot" option is particularly difficult, because rot seems to be a somewhat individualistic phenomenon, often arising from specific local details, or from gross inattention. There are however a number of strong clues in the location and especially the patterns of stains and rot, which help point to probable causes. Whilst any well-defined pattern is likely to be a pointer to a particular cause, the patterns DAMP looks for are geometric patterns (particularly those corresponding to the construction elements of the building, such as the location of studs), and whether the stains change with cold, with wet, or with time.

In Figure 2 is a representation of the decision tree used in preparing the "rot" option of DAMP. This is chosen as it is the simplest of the three options. Although very cryptic, Figure 2 should give some idea of how the process functions, with single, multiple, or zero conclusions being offered in various circumstances. Notice that even where a firm conclusion is indicated, there is still no detail on exactly where or how the moisture has entered. That information, plus confirmation of the diagnosis, must be supplied by direct inspection of the particular site.

4. Conclusions

A description of the difficulties confronting human experts attempting to transform their expertise from an associated or intuitive form, into a rule-based form, has been outlined using the DAMP Knowledge Base as an example. The DAMP Knowledge Base is also briefly described.

These difficulties include:-

a. the two forms of knowledge differ too widely

b. the "intuitive" form of existing human knowledge is difficult to access, and cannot be transferred. Only somewhat stereotyped representation of that knowledge can be transferred.

c. the physical models used by human experts to aid or test their diagnoses are not (yet?) representable in rule-based systems.

d. pattern recognition is a key item to human experts, but is not (yet?) available to rule-based systems.

e. limited ability of rule-based systems to detect false answers, which although usually unintended, can be deliberate or malicious.

f. rule-based systems cannot readily include general knowledge of the real world, although commonly important.

g. uncertainty of how users will interpret unsupervised questions and diagnostic conclusions.

Acknowledgements:

The assistance of co-workers involved in diagnostics in other technologies is gratefully acknowledged, particularly W. Sharman, R. Bishop, H. Brown and A. Bennett of BRANZ, and N. Brown of C.S.I.R.O.

Bibliography

Buis, M. Hamer, J. Hosking, J.G., and Mugridge, W.B. An Expert Advisory System for a Fire Safety Code. Contract Report 35-30-002 to BRANZ, 1986.

Dechapunya, A.H. & and Trethowen, H.A. Development of DAMP Expert Systems, to be published.

Sachdeva, P. DAMP - a diagnostic system for architectural moisture damage problems. Aust. Computer Journal Vol 17 No. 1. Feb 1985 p 27-32

Factor	"Intuitive-Based" System (Human)	"Rule-Based" System (Computer)
Availability to date	available, (but few good experts)	not available
Reliability of diagnosis	high	variable
Limits (to extent of accessible knowledge)	unlimited	rigid limit
Flexibility (to cope with unexpected variations)	high	zero
Consistency of diagnosis	usually consistent	totally consistent
Recordability of case details	records often imperfect	records can be exact
User training needs	very lengthy	minor
Form of knowledge	associative	heirarchal

TABLE I. COMPARISON OF TWO FORMS OF THE "DAMP" KNOWLEDGE BASE

Building Types:-

1. Houses
2. Multi-apartment
3. Office
4. School
5. Dry Process Manufacture
6. Wet Process Manufacture
7. Warehouse
8. Cold Stores
9. Swimming Pool Hall
10. Sports Hall
11. Laundry

Symptom Types:-

1. Mould
2. Visible water
3. Water stains
4. Other stains - Tar Stains
   - Pattern Stains
5. Paint Blistering
6. Rot
7. Corrosion

Schedule of Available causes:-

Moisture Sources

Construction moisture
Ground water
Rain leaks
People (breathing washing cooking)
Process moisture
Pipe leaks

Transfer Mechanisms

Air movement
Seepage/capillary
Sunshine/temperature changes
Condensation
Vapour diffusion
Storage

TABLE 2. SOURCES AND SYMPTOMS OF BUILDING MOISTURE

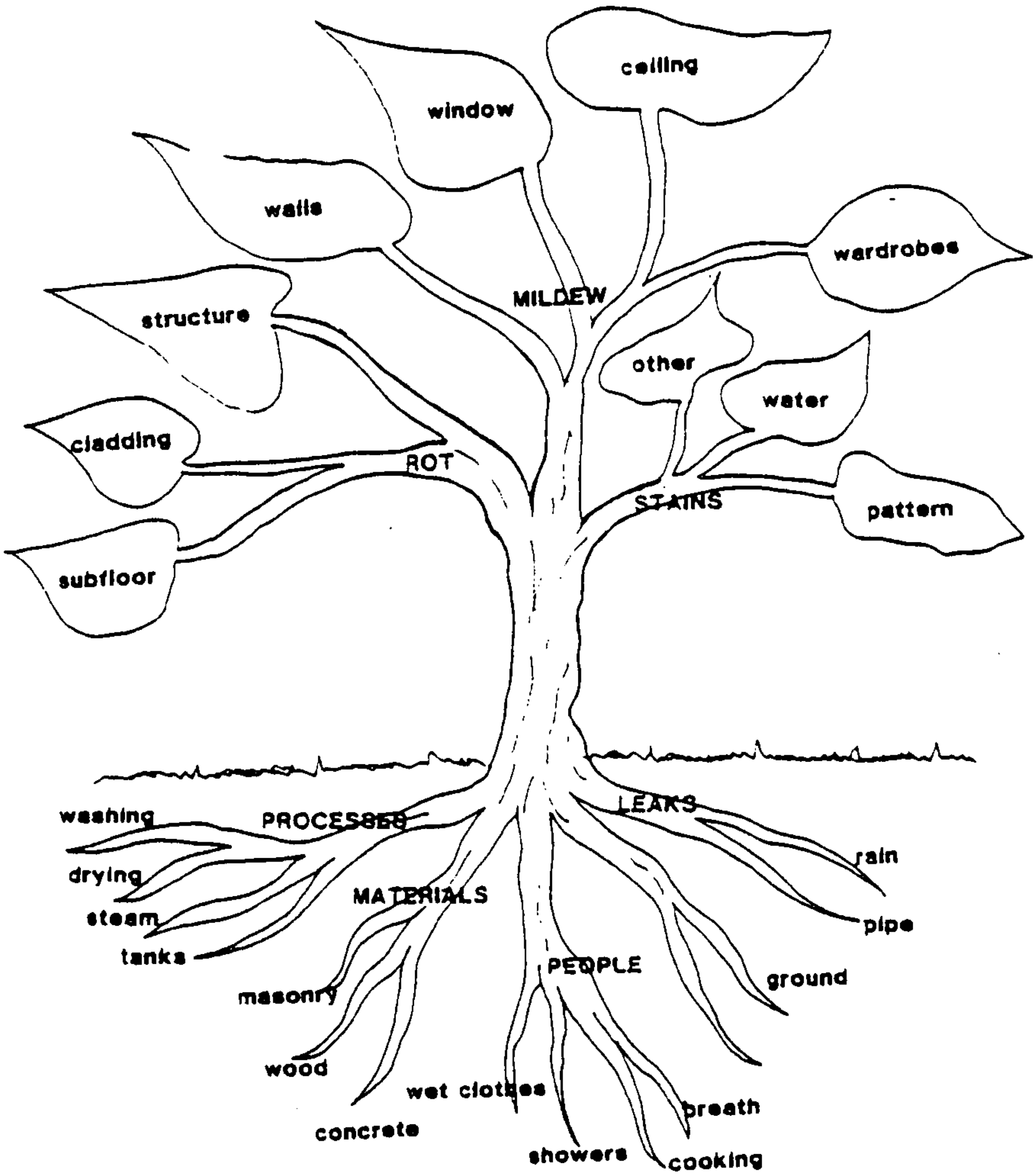
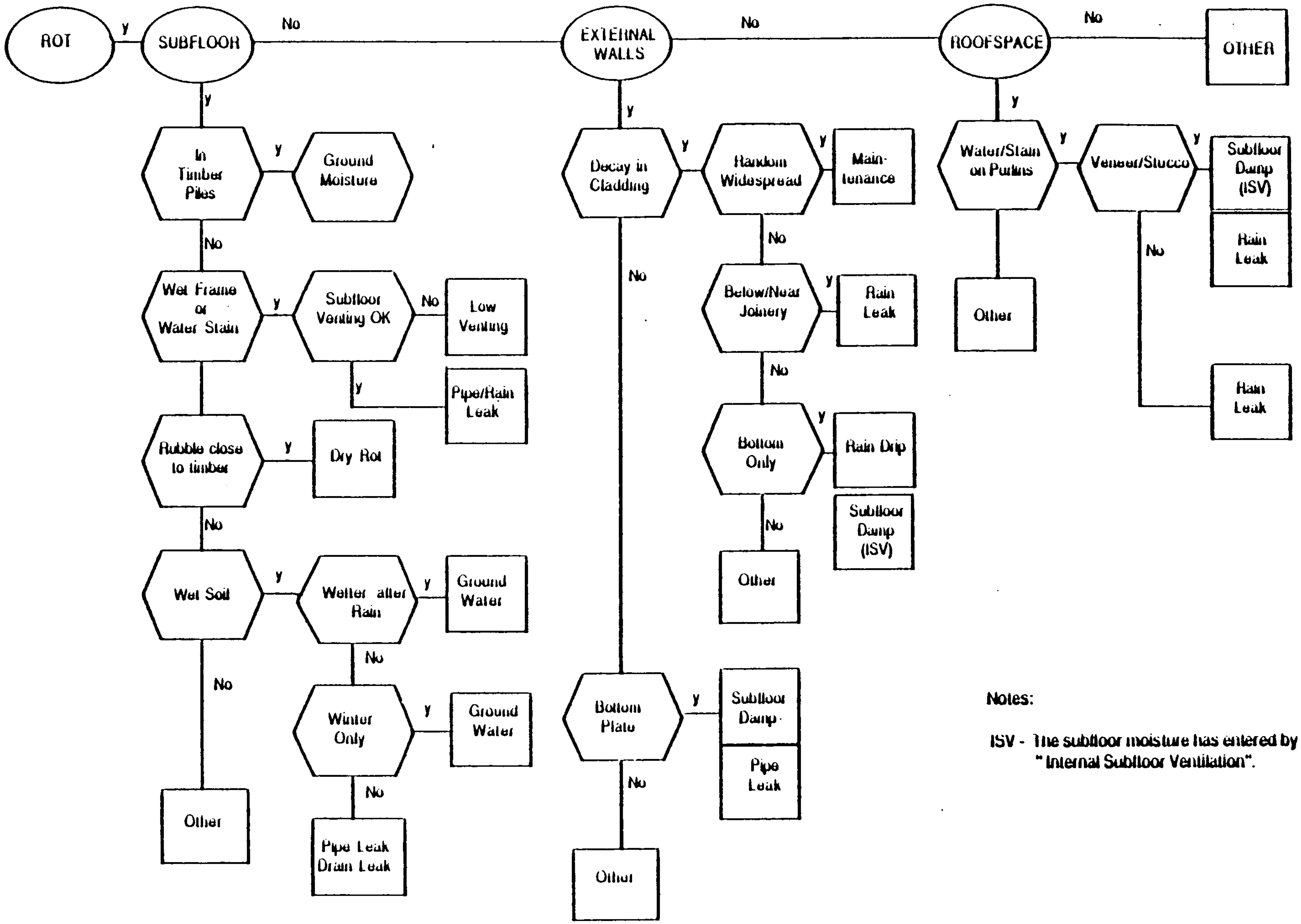


FIG. 1 THE DIAGNOSTIC TREE



Notes:

ISV - The subfloor moisture has entered by "Internal Subfloor Ventilation".

FIG. 2. Summary of present form of "Rot" option of "Damp" diagnostic rules

BUILDING RESEARCH ASSOCIATION OF NEW ZEALAND INC.
HEAD OFFICE AND LIBRARY, MOONSHINE ROAD, JUDGEFORD.

The Building Research Association of New Zealand is an industry-backed, independent research and testing organisation set up to acquire, apply and distribute knowledge about building which will benefit the industry and through it the community at large.

Postal Address BRANZ, Private Bag, Porirua

branz

ISSN 0111–7459