OPTIMIZING TELIDON TREE STRUCTURES

MAXIMISER LES STRUCTURES ARBORESCENTES DE TELIDON

Ann H. Schabas
Faculty of Library Science
University of Toronto
140 St. George Street
Toronto, Ontario M5S 1A1

ABSTRACT

Work was carried out under contract for the Canadian Department of Communication (DOC) in the Spring of 1980 to examine the Telidon tree structure set up by DOC and to restructure it as necessary to improve access to the pages of information. Guidelines were developed for optimizing the structure and a discussion of these guidelines will form the substance of the report to CAIS.

RESUME

Suite à une entente contractuelle, une étude a été menée au nom du Ministère fédéral des communications au printemps 1980 afin de vérifier la structure arborescente de Télidon établie par le ministère et de le restructurer si nécessaire afin d'améliorer l'accès aux pages d'information. Des directives pour en maximiser la structure ont été proposées et notre exposé présenté à l'ACSI portera sur la présentation de ces directives.

OPTIMIZING TELIDON TREE STRUCTURES

INTRODUCTION

This paper deals with features of Telidon tree structure design which I believe to be important for effective retrieval. A general familiarity with Telidon structure is assumed: its series of successive menus providing options to the user for document selection. I would like to preface my discussion by noting that, individually, each feature I will raise appears to be nothing more than good common sense, and focussing on them borders on labouring the obvious. On the other hand, they were not consistently applied in the original DOC Telidon data base and retrieval from it is short of satisfactory.

One of the constraints of a study of this sort is that we don't know our Telidon public. It has no precedent. We do know that we can't expect the same tolerance from Telidon users as from the users of on-line bibliographic data bases who, as highly motivated specialist information seekers, are generally prepared to invest time and money and effort into their searches, to tolerate considerable frustration in the process, and to put up with a mix of relevant and non-relevant 'hits' often heavily weighted to the non-relevant. The Telidon user picture will, undoubtedly, be very different. We can't afford to leave any stone unturned in providing mechanisms for quick, easy and highly satisfactory matching of Telidon user requests and 'documents', most of the time.

My thinking on these concerns is the outcome of a study carried out in the Spring of 1980 for the Behavioural Research Group of Communications Canada. Specifically this study dealt with DOC's trial Telidon data base, examining it for access problems and devising improvements to overcome these problems as much as possible.

The trial data base configuration consisted of a tree with approximately 800 branch endings. Only 200 of these branch endings actually held documents. The other approximately 600 were identified by penultimate menu options only. The nodes of the tree structure leading to these documents and document spaces were represented by menus which revealed the bases for subdivision along the search routes. The number of options at any one node was restricted to nine or fewer because of the nine-digit constraint of the communicating keypad, which in turn springs from our civilization's decimal number system.

I was directed to consider the set of documents and document spaces as the document set of one information provider. Its appropriateness to represent some future operational Telidon data base was not questioned. It was more a prototype (primitive form) than a model (standard). Nonetheless, I believe it was adequate for the problem identification objective of the study.

OPTIMIZING TELIDON TREE STRUCTURES

THE MAJOR AREAS OF CONCERN

There are five basic questions to be answered. What major categories should be chosen to partition the data base? If more than one menu level is involved, what order of precedence is most appropriate? How should options be expressed on menus? How should options be sequenced on menus? What auxiliary devices are available to supplement access provided by the rigid tree structure?

Discussion of these five questions will provide the main substance of this paper, and will be considered in turn. Reference will be made from time to time to the limiting size of the Telidon frame, which, in round figures is 40 characters by 20 lines. This size limit affects the number of options which can be presented, the way in which they can be expressed and the incorporation of auxiliary devices.

DATA BASE PARTITIONS

Any rigid tree structure results in a linear array of all documents, not unlike the shelves of books in a library arranged by some hierarchical classification scheme. The job of setting up such a scheme starts with the designation of a certain number of major categories in which it is expected all documents, present and future, will find a place. To expect to accomplish this without considerable compromise is unrealistic. No matter what basis of subdivision is chosen, authors of books or creators of Telidon documents will prepare items which cut across the preestablished boundaries.

Accordingly, the choice of categories must be made to accommodate most, but by no means all, situations. In the absence of a real Telidon user group from whom to derive preferential categories, I suggest three guidelines: that categories be chosen on the basis of their amenability to being expressed as mutually exclusive; that they be chosen to group together documents that are apt to be used together (e.g. advice on buying and for sale information pertaining to a type of merchandise); and that they be chosen to give emphasis to substantives rather than to actions, activities or attitudes. These three criteria function independently, but it is not anticipated that they will often be in conflict. Since there is an interdependence between these points and points to follow in considering the other questions posed, I plan to withhold examples until later in my presentation.

The artificial 'divide by nine' constraint (for Melvil Dewey it was 'divide by ten') of our number system must not be a factor in the choice of categories at any level. There are a number of ways to rise above it (no pun intended). The only

limiting factor is the frame size. I suggest that the number of options be set from logical considerations only, and then ways be sought to accommodate them. I propose three methods, which are all variations on the same theme.

Where the number is only slightly above nine, a pair or several pairs can be assigned the same number(s) e.g.:

5 EDUCATION: Opportunities, Resources

5 JOBS, Employment agencies, Careers.

An option can serve several specific categories, each of which is spelled out in an extension of some catchall umbrella term, e.g.:

4 MISCELLANY: Metric conversion tables; Horoscope

If there are many 'species' of a common genus they can be grouped in some way (alphabetically?) under option numbers, e.g.:

1 Anemia
Appendicitis
Arthritis
Asthma

2 Diabetes
Epilepsy
Hernias

3 Influenza
Kidney disease
Liver disease

4 Pneumonia
Psoriasis
Stroke
Tuberculosis

5 Ulcers
etc.

A fourth method of handling the problem is to interpose groupings on an intermediate menu. But this is strongly discouraged. The category names of these intermediate groupings are often arrived at artificially and can reflect overlap which is very misleading to the user faced with a menu of options which are not mutually exclusive.

In all three recommended methods, the duplicate postings must be resolved by the next menu. There is a gain in logical categorization without inappropriate and ambiguous intermediate grouping, at the expense of an additional menu. It is my contention that the user is better served by more (slightly) but clearer (unambiguous) choices.

CATEGORIZATION ORDER

Some types of material lend themselves to alternative approaches. Some users will approach restaurants by cuisine type, others by neighbourhood. For used cars, does age (year) take precedence over make? For real estate, does price preempt

district? In all cases it depends on the user. The only way, within the rigid tree structure, to provide both approaches where there is a good case for both is to duplicate the documents. For economic reasons this may well not be viable. Perhaps the document creator (the advertiser) will be given the choice of one or other or, for a price, both.

Where a choice must be made in the design of the tree, the decision to sequence option levels of this type should be made by persons knowledgeable in the field and after audience testing. The problem is well known to librarians where it is referred to as facet ordering. Whatever the decision, it must be recognized that at least one group of users, hopefully a small one, will be disserviced. This kind of limitation makes alternative approaches via cross references and keywords attractive (more on this later).

OPTION WORDING

Options must be worded very carefully. The user must be very sure that he is choosing a route that will get him to the information he wants. This is particularly true at the high levels of the tree -- in the first menus encountered. There is obviously a severe space limitation on wording and there is also the recognized aesthetic advantage of keeping text to a minimum. But skillful menu designers should pack as much information into the menu options as possible.

At the lower levels, further down the branches, options can often be quite brief because the context is fully established by the hierarchy of options already chosen. But unless the essence of this hierarchy is carried over there is the danger of the user losing it. My first recommendation in this regard is that all menu frames have titles. A menu frame title serves as a 'security blanket' to guard against misunderstanding. The further down the tree, the more important its role.

The wording of an option on a menu is context-dependent in another way. It must be examined vis-a-vis the other options on the same menu to ensure that any ambiguities and overlaps are resolved. Consider, for example, a menu which includes the following two options:

1. News, Weather, Sports

4. Home and Community: Entertainment, Hobbies, Sports

A change to:

1. News, Weather, Sports news

4. Home and Community: Entertainment, Hobbies, Sports activities

will prevent many a wrong choice regarding sports. Once a conscious effort is made to look for potential confusions between options, it is surprising how many are found and how easy they are to repair. And if a given situation is not easily resolvable it suggests a poor choice of categories in the first instance. The solution in such cases may lie in some restructuring of the tree.

Graphic design is not an area of my expertise, but as frames carry more text for reasons of clarity perhaps some of the resulting clutter can be offset by imaginative use of typographies, colour and blinking devices.

OPTION SEQUENCE

The importance of option sequence should not be overlooked. Users tend to scan options from top to bottom. The inclusion intended at one option can be assumed set, to some extent, by options that precede it in the list. This is most obvious when a list of specific options is followed by one which simply says 'Other'. But also consider the following:

6. Babies' and children's needs: Clothing, Furnishings
7. Clothing
8. Furnishings

Clearly, options 7 and 8 exclude babies' and children's clothing and furnishings, respectively. And the user can be expected to infer this from the sequence of the options.

A word of caution is in order regarding the use of the catchall option 'Other', and its close cousin 'Miscellaneous'. My first advice is don't use them unless there is no better alternative; and I can't conjure up in my mind a situation where there is no better alternative. The menu frame designer can always do better. He/she knows what is down the trail and should find some brief way of saying what it is.

The presence of an option like 'Other' has, I believe, a disconcerting affect on the user. A few may enjoy the suspense, but more will be annoyed by the lack of information. The technique has the flavour of those bins of 'mystery packages' used by drug stores to get rid of left over merchandise.

Consider another sequencing problem:

1. Biographies
2. Novels
3. Fiction

4. Science-Fiction
5. Paperbacks.

Does the presence of the option 'Paperbacks' mean that the preceding four options are hardbacks? If this inference is correct, it is only fair to the user to put 'Paperbacks' first, and list the other menu options under 'Hardback'. Likewise, if 'Novels' and 'Fiction' exclude 'Science-Fiction', this would be clearer if 'Science-Fiction' came before the other two. And the distinction between 'Novels' and 'Fiction' is by no means clear. Some qualification is in order, either by extending the text of the options or by merging them as one option and making some distinction at the next menu.

AUXILIARY ACCESS DEVICES

The strictest attention to these optimizing criteria still leaves problems for the rigid tree structure. The user whose approach differs from the facet order chosen by the tree designer will not be well served; and the documents which cut across category boundaries will be missed or hard to find by a large group of users. Both problems can be overcome by duplicating documents but this would greatly inflate the data base. The user will, eventually, learn the system after repeated use, but will his patience survive the learning period? I also believe that it will be harder to learn and retain the organization of this volatile, on-line tool than it is to learn to find one's way around in something more concrete, such as a reference book, a newspaper or a supermarket, where one can always orient one's present position in the context of the whole. More on this 'sense of location' in the next section.

The introduction of a syndetic structure will go a long way towards alleviating the need to duplicate documents. (It may, on the other hand, add to the 'sense of location' problem.)

Syndetic structure means cross references. These can be of the form 'For Car rentals key xxxx' to take the user from the menu offering car purchase and leasing alternatives to a menu or document on local transportation facilities. A line or two at the bottom of a menu could easily accommodate such pointers to other branches. Alternatively, with some software enhancement, such links could be in the form of regular one digit options, and, consequently, transparent to the user.

Cross referencing opens up another access device: the on-line directory. The user would choose in succession from a series of narrower and narrower alphabetic ranges until he reached a sought for segment of the alphabetic index of topics. Here he would be directed to key the precise frame number -- assuming he chose a term present in the index. There are synonym problems with this mode of access and it is somewhat cumbersome

even when it does prove fruitful, but viewed as one of several complementary access devices its value should not be underrated.

It should be noted that, powerful as cross referencing is, it does not help the 'cuisine type/neighbourhood dichotomy for the restaurant seekers' problem described earlier.

SENSE OF LOCATION

The problem of providing a 'sense of location' within the data base was mentioned in the last section. It is a problem which has been identified and felt by videotex users, but no measure of its seriousness has yet been ascertained. How important is it to a user to know how one particular frame fits into a broader schema and to be able to move back and forth among a set of related frames? These are givens in the world of books and the like. They are not givens in Telidon-like tools.

I personally feel that these concerns are real enough to take what steps we can to enhance the 'sense of location'. One way is to provide context-setting titles to frames wherever possible: on all menus and on documents as long as title presence does not impinge on the graphic quality. And all menus and documents should carry their frame numbers, in some predictable position.

Ample cross references of the kind already described will also help. Cross references from documents to some carefully selected higher node in the same branch would also be most helpful. These were included to some extent in the DOC tree, for example: on a particular restaurant listing the message 'To return to Cuisine Index key xxx'.

CONCLUSION

This has been a brief overview of problems identified following a careful study of a particular videotex data base. Techniques for minimizing the problems have been suggested. Their effectiveness awaits formal testing.

If one subscribes to the view that optimum access will ultimately involve the complementary application of a variety of access modes, then this report only deals with part of the picture. Considerations of keyword access and Boolean logic open up a large new area for investigation. Work to develop and assess keyword indexing and post-coordinate retrieval for Telidon is underway.

BIBLIOGRAPHY

Ball, A.J.S. "Videotex: Chimera or Dream Machine", in Canadian Library Journal. Vol. 38 (February 1981), pp. 11-15.

Bown, H.; et al. "Telidon Videotex and User-Related Issues", in Behavioural Research on TELIDON I. Ottawa, Department of Communications, 1980. pp. 9-25.

Lane, Martin. "Videotex Database Structures and Cross-Referencing." 1980. 2 p. (Photocopy)

Lee, Eric and Latremouille, Susane. "Evaluation of Tree Structured Organization of Information on Telidon", in Behavioural Research on TELIDON I. Ottawa, Department of Communications, 1980. pp. 231-242.

Phillips, Dorothy. "The Implications of Telidon for Information Science", in Behavioural Research on TELIDON I. Ottawa, Department of Communications, 1980. pp. 27-39.

Thomas, Hilary B. "Tree Structures: the Root of Videotex?" 1979. 9 p. (Photocopy)

Williamson, Nancy J. "Viewdata Systems: Designing a Database for Effective User Access." Canadian Journal of Information Science. Vol. 6 (1981). (Accepted for publication)