Claus Ritzler

Comparative Study of
PC-supported Thesaurus Software

This article presents the results of a comparative study of three PC supported software packages (INDEX, PROTERM and TMS) for development, construction and management of thesauri and other word material with special regard to hardware and software requirements, handling and user interface, and functionality and reliability. Advantages and disadvantages are discussed. The result shows that all three software products comply with the minimum standards of a thesaurus software. After inclusion of additional features distinct differences become visible.

(Author)

1. Background

This article presents some results of a comparative study (1) made a year ago as part of the author's thesis.

The essential part of this work was the comparison of three PC-supported software packages for development, construction and management of thesauri and other word material with special regard to

– Hard- and Software requirements
– Handling and user interface
– Functionality and reliability

The author has drawn up 8 different tables of parameters to evaluate the software.

The 8 tables and their headings had been as follows:
– Software producer information
– Hardware and software requirements
– Software installation
– Operation and user interface
– Functionality: Input of terms and classifications
– Functionality: Creation of term relationships
– Batch functions
 (involving e.g. different input and output functions)
– Management and system functions
 (involving e.g. user access control, reorganisation of the data etc.)

With this catalogue of evaluation criteria it was possible to identify the differences between the products providing a basis for discussion of the salient pros and cons. Additionally it was possible to determine whether all three software products comply with the minimum standards of a thesaurus software and under which conditions these products are suitable for establishment, development and maintenance of individual thesauri.

2. The different thesaurus software packages and their producers

Three products have been compared: one British and two German thesaurus software programs.

The British one is called TMS (Thesaurus Maintenance System) Version 1.0 and was developed by Pyramid in Reading. The TMS thesaurus software package is used, mainly in commercial libraries and information centres, for constructing and maintaining thesauri which will be used eventually for indexing company documents. The user can purchase this product as part of the Computer Aided Library Management System (CALMS) or as a standalone thesaurus construction system. TMS is priced at 1.840 DM.

One of the German products is PROTERM-T Version 2.5. It is a software package from PROGRIS in Berlin. The software PROTERM-T was especially designed for constructing and maintaining vocabularies. PROTERM-T is a standalone system, but it is possible for the user to buy up to 5 additional modules. These modules are not included in the comparison. The price for the main software PROTERM-T is 1.200 DM.

The other German package is named INDEX Version 4.1. It is a product of ERNST LUKAS. INDEX is a universal package for developing and maintaining thesauri, vocabularies and classifications. For INDEX are also up to 8 modules available, which are not included in this comparison. With INDEX the user receives a Dataflex Runtime License or a Dataflex Development License from Data Access. The costs for INDEX are 6.200 DM plus the Dataflex Runtime License, which is approximately 300 to 2.000 DM.

3. Software and Hardware requirements

PROTERM-T and TMS run under MS-DOS 2.1 upwards. INDEX requires MS-DOS 3.3 or another Operating system such as PC-DOS, UNIX/XENIX, OS/2 and others. All three programs can be installed on a PC-XT or AT, with a hard disk and 512 KB memory. For INDEX the producer recommends an Enhanced Graphics Adapter with a suitable display. TMS requires one disk drive to be free for the program disk, because it is not possible to copy the program to the hard disk. Thesaurus data, however are saved on the hard disk. TMS also needs a printer, because it is not possible to display all kinds of reviews on the monitor.

1. Software Producer Information		INDEX Version 4.1	PROTERM-T Version 2.5	TMS Version 1.0
AwA	Software Producer Address	ERNST LUKAS Leipziger Ring 16 6054 Rodgau 3	PROGRIS GMBH Auguste-Viktoria-Str. 64 1000 Berlin 33	PYRAMID SOFTWARE PRODUCTS LTD. 9 Church Street Reading RG1 2SB United Kingdom
AwA	Status of the Producer	Subcontractor	established Company	established Company
AwA	Company Foundation	1984	1977	1982
AwA	Number of Employees altogether	1 Employee	8 Employees	7 Employees
AwA	Custom Services	Consultation Course of instruction	Consultation	Consultation Course of instruction
AwA	Number of References	60 Single-User Installations 5 Multi-User Installations	24 Installations	40 Installations

AwA signifies: Generally important product information. These represent the necessary requirements of a given product.

Table 1: Software producer information

2. Hardware and Software Requirements		INDEX Version 4.1	PROTERM-T Version 2.5	TMS Version 1.0
AwA	Hardware Basic Equipment	PC-XT, PC-AT IBM kompatible PS/2-Series and others 512 KB Memory Hard disk Enhanced Graphics Adapter with a suitable display recommended	PC-XT, PC-AT IBM kompatible 512 KB Memory Hard disk	PC-XT, PC-AT IBM kompatible 512 KB Memory Hard disk Minimum 1 Disk drive Printer
AwA	Operating System	MS-DOS Version 3.3	MS-DOS V. 2.0 upwards	MS-DOS V. 2.1 upwards
AwA	On other Operating Systems usable	e.g. PC-DOS, UNIX/XENIX, OS/2 and others	No	No
AwA	Standard-Software Basic Equipment	INDEX Dataflex Runtime License or Dataflex Development License	PROTERM-T	TMS
AwA	Number of additional Standard Modules	8 Modules	5 Modules	No
AwA	Costs of the Single-User Basic Version	DM 6.200	DM 1.200	DM 1.840
AwA	Costs of Licenses for other Software	DM 300 to DM 2.000	-	-
AwA	Costs for additional Standard Modules	DM 150 to DM 3.320	DM 190 to DM 380	-
*	Capable for Multiple Places	Yes, on all common PC-Multi-User-Operation-Systems	No	No
*	Capable for Networks	Yes, on all common Network-Operation-Systems	No	No
AwA	Development System used	Dataflex	Clipper	Turbo Pascal
AwA	Data Structures	Semantic Network	Simulated Relational Model	Relational Model with purely hierarchical structured elements
*	Source Code available	Matter of Negotiation	No	No
AwA	Quantity of Delivery	3 5¼ inch Diskettes á 1,2 MB 230 Pages Manual Basic Set of Test Data	2 5¼ inch Diskettes á 360KB 100 Pages User-Manual	2 5¼ inch Diskettes á 360KB 27 Pages Operating Guide
AwA	Documentation and Literature - Quality - Languages	Manual Very good German	User-Manual Good German	User Introduction Satisfactory English

AwA signifies: Generally important product information. These represent the necessary requirements of a given product.

* signifies: Here are features required in application cases or by the system's environment.

Table 2: Hardware and software requirements

4. How are the programs presented to the user?

The initial presentation of a software product is comparable with the first interview between a personal manager of a company and the person who is looking for a job. The first impression must be good.

The first menus the user gets are in all three programs similar. One is more detailed the other less.

In Fig. 1 with the main menu of TMS there are mainly two parts called construction and review. On the construction side we see the input components like the options to enter terms, relationships and scope notes, the option to change or delete terms, the option to add facets or even the option Re-index what is normally a management or systems function. On the other side the user will see more or less all output functions.

In Fig. 2 with the main menu of INDEX we see a clear menu with 7 options, which contains completely different functions. The normal user needs option 1, to carry out all necessary operations for developing his thesaurus, classification or source register and naturally option 2, with the data output functions. The other functions are more or less for the system manager, they involve features like reorganisation, data backup, the definition of relationships and control parameters etc.. The background is coloured in blue, the line 1 is coloured in grey. All colours are optional and may be changed by the user.

In Fig. 3 with the main menu of PROTERM we have three areas as well. First the dialog functions to enter, delete and list terms, than the batch functions for reorganisation and output. The last bit contains the different management function.

However all three thesaurus programs are presented in a different way to the user.

INDEX works with different types of objects. First there are forms, these forms are structured into different fields and look like record cards. And there are lists. See Fig. 4.

INDEX provides all kinds of interactive possibilities; the screen layout at the different working levels is fully coloured. In spite of the complexity and the different functions, INDEX is completely driven by using the ten function keys F1 to F10 and the cursor keys. Additionally the user gets a menu line on the upper part of the screen by using the function key F10. This main menu line with various submenus is partly equivalent to the function keys. At some stages, especially at the systems management level the user is presented with pull down menus. All these interactive possibilities and the different usage of colours are unified at every stage of the program. Each colour has its own meaning, aiding the user in learning and understanding the software and its working processes.

PROTERM also has menus at different working levels. The options for each level are numbered upwards. The monochrome display of PROTERM and TMS operate in conjunction.

With

— different cursor keys, or

— by entering a number, or

— by entering different key combinations

the user can choose different options from each menu. The interactive possibilities are not unified in PROTERM. If the user wish to jump to another (e.g. a two or more levels higher) menu, he must path through all intermediate menus sequentially. It is not possible to jump between the different levels as it is in INDEX because the function keys in PROTERM are not used at all and only for statistical operations in TMS; the operation is cumbersome.

TMS is not menu-driven, it is driven by commands, cursor keys or partly by function keys. Alternative interactions are not possible.

Unlike INDEX, PROTERM and TMS have no help facilities.

5. Input of terms

This part describes the input and search facilities of the different programs.

The data entry mode in TMS is accessed by typing the number 10. In PROTERM and INDEX it is selected by choosing option 1 from the main menu. The user is presented with a new display layout, which is different for each program.

In TMS the user may enter a term by typing the specific term or by typing the number code for the term, if known. Terms are limited to a maximum of 40 alphanumeric characters only.

The word can be described with a scope note or a facet. To register a scope note the user has to type the word again, after pressing the function key F3. A field with maximum 1.110 character space appears. To locate a facet the user has to go back to the main menu to activate option 12; making operations inconvenient. The facet can be 20 characters long. The facet feature is primarily intended to relate to a classification scheme. However, as yet, no facilities exists for sorting and printing the thesaurus by the facet.

In PROTERM data entry, data display, data correction and data deletion are all available at one level, but it does not include term correction or term deletion. This is included in option 2 of the main menu. Each term can be 60 characters long. The user has the possibility to enter a scope note with a maximum of 150 characters and a notation with 6 characters. The program proofs each term simultaneously to the data entry. A small display shows a changing term with the same sequence of characters as inserted. The program also displays a list of terms by pressing function key F2. The user can choose words from this list.

It is not possible to insert terms successively. The user always gets the menus for the different relationships first. Interrupting this process is possible by typing the "Hatchmark"(#) and some carriage returns. This is again a very time consuming procedure for users who want to enter lists of terms before they start with the relations. Especially at the beginning stage of a thesaurus project, where the word material will be collected first. It is not possible to enter short terms who are identical with the first characters of already existing terms. In the

Thesaurus development

Construction	Review

Construction

10. Terms, relationships and scope notes
11. Change or delete terms
12. Facets
13. Language alternatives
14. Indentation levels

20. Re - index

Review

50. View thesaurus

60. Term list

62. Single level term list
63. Single term systematic list
64. Alphabetic systematic list
65. Alphabetic NT1, NT2 list
70. Create language index

80. Change print parameters

Please enter your choice :

Fig. 1: The main Menu of TMS

```
AXEL SPRINGER VERLAG - 2000 HAMBURG        109867 30/08/90 10: 8   279312
```

```
  INDEX 4.1

1 INDEX-Datenbank
2 Datenausgabe
3 Definitionen
4 Systemverwaltung
5 Dateiverwaltung
6 Datensicherung
7 Aktuelle Datenbank wechseln
```
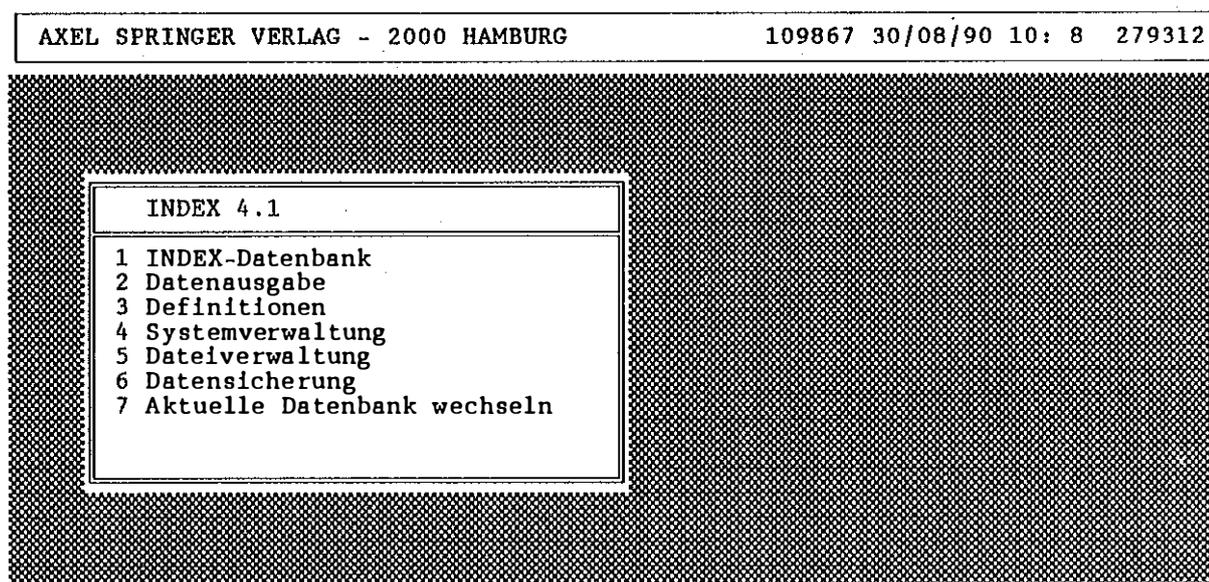
Fig. 2: The main menu of INDEX

PROTERM-T Vers. 2.4   (c) PROGRIS 1987, 1988   Dienstag, 07.08.90

```
HAUPTMENU

  Dialog-Funktionen
  1 > Beziehungen anzeigen/erfassen/löschen
  2 > Terme korrigieren/löschen (Termnummer erforderlich)
  3 > Term-Listen anzeigen

  Batch-Funktionen
  4 > Term-Listen drucken
  5 > Reorganisation: Löschen/Sortieren/Sichern

  Verwaltungs-Funktionen
  6 > Grundstruktur anzeigen
  7 > Erfassungs-/Anzeige-Formate anzeigen/ändern
  8 > Druckformate anzeigen/ändern/erstellen


  0 > ENDE
                                                  Ser.No.
```
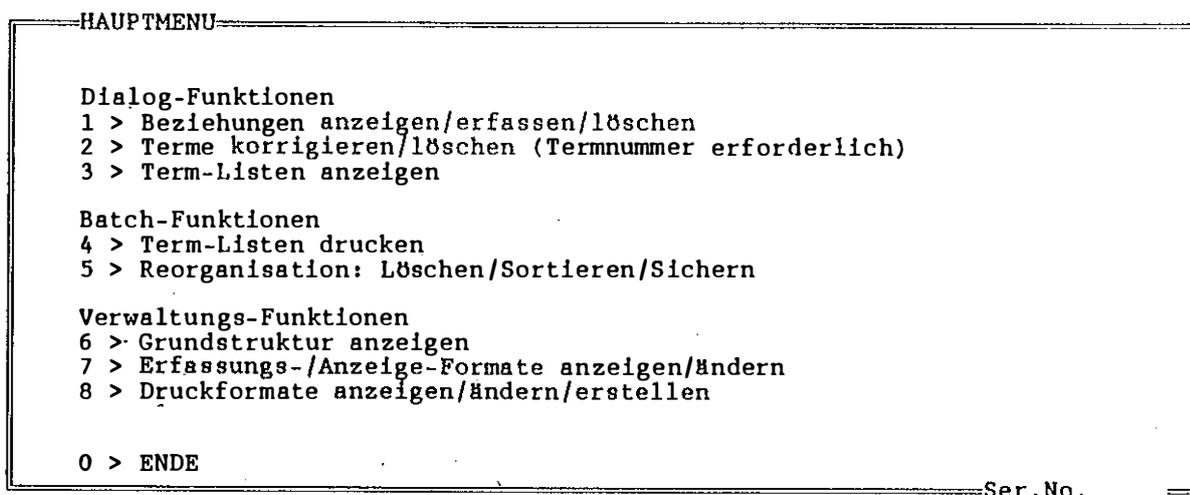
Mit Cursor auswählen, dann <ENTER> oder Ziffer wählen

Fig. 3: The main Menu of PROTERM

```
┌─────────────────────────────────────────────────────────────────────────┐
│  Klassifikation nach Notation                                             │
│  ┌──────────────────────────────────────────────────────────────────────┤
│  . * Aachener und München Beteiligungs-AG allgemein. AfAa0010   ::FA      │
│  . * Aachener Rückversicherungsgesellschaft allgemein. AfAa0015  ::FA     │
│  . * abc-Autorenbuchhandlung allgemein................ AfAa0020  ::FA     │
│  . * Ackermann-Göggingen allgemein.................... AfAa0025  ::FA     │
│  . * A┌───────────────────────────────────────────────────┐0030   ::FA   │
│  . * A│ KLASSIFIKATION              Nr:   5231  S:  M:     │0035   ::FA   │
│  . * A├───────────────────────────────────────────────────┤0040   ::FA   │
│  . * A│ Notation              │ B T  Eb U │ 1 2 3 │ F │    │0045   ::FA   │
│  . * A│ AfAa0005    ::FA      │ K    1    │       │   │    │0050   ::FA   │
│  . * A├───────────────────────┴───────────┴───────┴───┘    │0055   ::FA   │
│  . * A│ . * Aachener und München Lebensversicherung AG     │0060   ::FA   │
│  . * A│ allgemein                                          │0065   ::FA   │
│  . * A└────────────────────────────────────────────────────70   ::FA     │
│  . * AG│ WORT                       Nr:   8428 H: 0 S:  M: │ 75   ::FA     │
│  . * AG├───────────────────────────────────────────────────┤80   ::FA     │
│  . * Ag│ Aachener und München Lebensversicherung AG        │85   ::FA     │
│  . * Ag├───────┬──────────┬───────────────┬─────┬─────┬───┬─┤90   ::FA     │
│  . * AG│ Sp    │ Sgr Fa   │ Wo G Gm Sch   │ 1 2 3 │ Z1 Z2 │ F │ T │93  ::FA│
│  . * AG│       │ In  J    │         P5    │ 6     │ D  X  │   │   │95  ::FA│
│  . * AG└───────┴──────────┴───────────────┴───────┴───────┴───┴───┘00 ::FA│
└─────────────────────────────────────────────────────────────────────────┘
 -> Objekt  Zoom  Verbinde     Liste  Suche  Form  Mark     Hilfe  Weiter  Ende
```

Fig. 4: The objects in INDEX; forms for words and classifications and a list in the back.

```
    Pyramid Software  0734 595633      *   *   *   *   *        10/8/1990
                                                          F4  F5  F6  F7  F8  F9
   Rl  No.    Term                                        US  UF  BT  NT  RT  SA
        53 Volkswagen                                      2   0   0   0   1   0
   SN

   US   54 VW                                              0   1   0   0   0   0
   US   55 VW Käfer                                        0   1   0   0   0   0
   RT   56 Audi                                            0   0   0   0   1   0
```

Fig. 5: Data entry mode in TMS

```
PROTERM-T Vers. 2.4   (c) PROGRIS 1987, 1988            Dienstag, 07.08.90

┌══ERFASSEN / ÄNDERN / LÖSCHEN══════════════════════════════════════════════┐
└═══════════════════════════════════════════════════════════════════════════┘

Bitte Term eingeben: a

                    D-Zug
                    Damenfahrrad
                    Dampflokomotive
                    Diesellokomotive
                    E-Lok
               --->Eilzug
                    Eisenbahn
                    Kfz
                    Kraftfahrzeug
                    Motor
                    Öffentliche Verkehrsmittel




Auswählen mit Cursor oder <PgUp> <PgDn> <Ende> <Home>
Term übernehmen mit <Enter>    Liste ausschalten mit <F2>
```

Fig. 6: Data entry mode of PROTERM with a list

German language we have often the problem with composita: e.g. Eisenbahnschaffner; Eisenbahn; Eisen. In this case the user has to enter the shorter terms first or he has to insert a blank at the end of the shorter term.

Homonyms and other words with identical character sequences cannot be entered in PROTERM and TMS. In INDEX a homonym term can be marked in a special field with an H. In the search mode of PROTERM no distinction is made between upper and lower case which can be very helpful.

In INDEX the user is presented with a blue formula to enter the words. This blue formula represent a 3by5 card from the card index (Fig.4). The formula is used for recording, searching and relating the data. You have different formulas for words, classifications and sources. In the formula for words you enter a new term and in the different 16 smaller fields you may add specific information about the term. Special information can be a language code, a subject code, a facet and others. In this case the user is completely free in his decisions, he can define the categories by himself. Some of the fields are indexed. This is a very helpful feature for sorting and listing the material for various aspects.

This is possible in PROTERM only partly by combining the term with a notation. But each notation may have only one combination with a term. So, if one wants to add more specific information one should contact PROGRIS, they provide an additional module for that.

Recorded with each word saved is the entering date, and the date of the last change is saved together with the name or code of the user who did the entry/change. The maximum size of a controlled term is 50 characters. The term itself can be as long as wanted, but it is not controlled any longer. That means, if one has two terms which differ at character 51, INDEX will not accept it.

During the recording of terms it is possible to define a master record for such terms which have e.g. the same descriptions. In this case one needs only to enter the term into a copy of the master containing the fixed information. The user has various possibilities to search, correct and delete terms. He can do this in a list or in the formula. It is no problem to switch between a list and a formula. In both types of objects he is able to browse forward and backwards in the term material to make further selections. The user can also mark words in a list for creating a separate working list. In INDEX scope notes have no limit in their length. To enter a scope note you have to open a text field. This mode is supported with wordprocessing features. They enable the user to delete and insert lines, to search for a word or a phrase and more. The user has the possibility to define different kinds of notes (e.g. Definitions, History Notes, Scope Notes etc.) by himself.

6.   Different relationships and consistency control

INDEX provides the user with the choice between 24 different relationships.

In INDEX it is possible for the user to create and define more relationships if required.TMS and PROTERM provide 5 different relationships to the user. In all three programs the minimum standard relations: Syn-

onymy, Hierarchical Relation and Associative Relation are involved.

All relationships have been designed to conform to British, German and international standards. The "See Also Relation" in TMS is identical with the relationship "Identical Use" in INDEX. As mentioned, it is possible to define more relationships in INDEX, but not in PROTERM or TMS. PROGRIS provides additional modules for additional relationships.

Furthermore all or parts of the control routines may be switched off in INDEX as if you want to ignore special routines. INDEX has completely user controlled consistency checks. In this part the user is completely free in his decisions, but it is worth to think about every change of the default values very carefully. INDEX provides more control routines than the other programs because a different description of·a term can be controlled as well. For example the user can forbid to relate terms of a subject A with terms of a subject B.

In all three programs the reverse relation is assigned automatically after saving a relation between two terms, but the control functions are not as strict as in INDEX, specially not in TMS. For instance: once assigned terms as non-descriptors can be related to all other terms, just like a descriptor. In fact everything can be related with everything. There is only one exception. A descriptor cannot be combined with itself. If so, the program will stop.

In PROTERM the consistency checks are done perfectly well. If you want to enter more than one relation to a non-descriptor, the term has to be assigned as polysem first, otherwise it is not possible.

Automatic reorganisation of very complex relationships after changing a descriptor to a non-descriptor will only be done in INDEX which is very comfortable. In the other two programs one has to delete the relationships first before assigning the non-descriptor to a descriptor, thereafter one has to build up the relationships again. This can be very time consuming, especially if the descriptor had a lot of relations on various levels.

In all three programs there are no limitations regarding the number of relationships per descriptor.

7.   How to enter the relationships

In PROTERM the user gets, as already mentioned, each kind of relationship after having entered the term. He can affect the order by typing the minus sign (to get the previous kind of relationship) or by typing an additional carriage return (CR), than he gets the next kind of relationship. The order and the abbreviation of the relationship can be changed by the user under the management functions. In the same mode, relationships per term can be displayed and deleted. If he wants to delete all relationships of a term at once, he has to delete the whole descriptor under menu point three and two.

In TMS it is similar to PROTERM, to enter a new relationship the user has to type the abbreviation for the relationship followed by CR, followed by the term or its number. Just as many relationships may be entered as are required. Whilst manipulating relationships of a certain type, it is convenient that the user can exclude other

```
┌─────────────────────────────────────────────────────────────────────────────┐
│  ┌──────────────────────────────┬──────────────────────────────────────────┐ │
│  │ Textfelder zu Wort           │ Text zu: Arabische Staaten        SCN   1 │ │
│  ├──────────────────────────────┼──────────────────────────────────────────┤ │
│  │ DEF  Definition              │ Staaten des arabischen Sprachraumes und der│ │
│  │ SCN  Scope Note              │ Arabischen Liga (ausgenommen der PLO), sowie│ │
│  │ HIS  History Note            │ dem Iran als nicht arabische Ausnahme am  │ │
│  │ ───  ───────────             │ Persischen Golf                           │ │
│  │ ───  ───────────             │                                           │ │
│  │ ───  ───────────             │                                           │ │
│  └──────────────────────────────┘                                           │ │
│       ┌──────────────────────────────────────────────────┐                   │
│       │ WORT                        Nr:   7186 H: 0 S:  M: │                   │
│       ├──────────────────────────────────────────────────┤                   │
│       │ Arabische Staaten                                  │                   │
│       ├─────┬────────┬──────────────┬───────┬───────┬───┬──┤                   │
│       │ Sp  │ Sgr Fa │ Wo G Gm Sch  │ 1 2 3 │ 21 22 │ F │T │                   │
│       │     │ Ge  S  │              │   0   │ D   f │   │  │                   │
│       └─────┴────────┴──────────────┴───────┴───────┴───┴──┘                   │
└─────────────────────────────────────────────────────────────────────────────┘
 -> Objekt  Zoom  Verbinde    Liste  Suche  Form  Mark    Hilfe  Weiter  Ende
```

Fig. 7: Scope note with its term in INDEX.

INDEX 4.1			PROTERM 2.5	TMS 1.0
1. Synonymy	Oescriptor	Non-Descriptor	X	X
2. Abbreviation	Full Descriptor	Abbreviation		
3. Identical Use	Term	Same Meaning		X
4. Quasisynonymy	Descriptor	Quasi-Synonym		
5. Alternative	Descriptor (alternative)	Same Meaning		
6. Combination	Single Descriptor	Combined Term	X	
7. Use	Basic Term	Example of Use		
8. Hierarchical Relation	Broader Term	Narrower Term	X	X
9. Generic Relation	Broader Term	Narrower Term		
10. Partitive Relation	Whole	Part		
11. Causality	Cause	Effect		
12. Appurtenance Relation	General Term	Instance		
13. Field Relation	Field	Element		
14. Associative Relation	Term	Related Term	X	X
15. Temporal Sequence	Predecessor	Successor		
16. Splitting	Predecessor	Successor		
17. Union	Predecessor	Successor		
18. Opposite	Term	Opposite Term		
19. Use Warning	Term	In Contrast To		
20. Production	Producer	Product		
21. Relation of Material	Material	Object		
22. Other Language	Language	Foreign Language		
23. Translation	Source Language	Target Language		
24. Systematic	Oescriptor	Notation	X?	X?
			one Notation/Facet per Term only	

Table 3: The different kinds of relationships in INDEX, PROTERM and TMS

relationships from the screen during editing. This can be done by using function key F4 to F9. These keys are used as toggle switches to turn on or off the display of each relationship type (see the right side of Fig. 5). To delete a relationship is quite easy. One has to move the cursor next to the relationship to be deleted and to press CRTL and D. It is not possible to delete more than one relation once.

The input of relationships in INDEX is different from the other two software products. Here the user gets two record cards on the screen after pressing two function keys. The contents of each card may be switched to one another, so that the upper content appears in the lower card and vice versa. After pressing a function key again a red clamp appears with the kind of relationship used last, in it. Now it is possible to choose another kind of

Wörter (G) alphabetisch	M	Sgr	Beziehungsarten
Öffentliches Verkehrsmittel			1 Synonymie
Öl			1.1 Abkürzung
Öltanker		WAS	1.2 Identische Verwendung
Paddel		WAS	1.3 Quasisynonymie
Paddelboot		WAS	1.4 Alternative

```
Pannen ┌─────────────────────────────────────────────────┐ ion
Parkh  │ Deskriptor          ]      Nr:     482 H: 0 S:   M: │ g
Parkp  │ ┌─────────────────────────────────────────────┐ │ e Hierarchie
Parks  │ │ Öffentliches Verkehrsmittel                 │ │ e Hierarchie
Parku  │ │                                             │ │ üt
Parkv  │ │ Sp │ Sgr Fa │ Wo G Gm Sch │ 1 2 3 │ Z1 Z2 │ F │ T │ gkeit
Passa  │ │    │     G  │             │       │       │   │   │ ehung
Passa  │ └─┌───────────────────────────────────────────┐─┘ haft
Passa  │   │ Nichtdeskriptor      │ 1   Synonymie  │ Asp:  │ olge
Pedal  │   │ Kraftwagen                            │
Pendler│   │                                       │
Personen│  │ Sp │ Sgr Fa │ Wo G Gm Sch │ 1 2 3 │ Z1 Z2 │ F │ T │ a
Personen│  │    │ STR G  │             │       │       │   │   │
Personen│  └───────────────────────────────────────────┘
Personen└───────────────────────────────────────────────────┘
```

Wort VERBINDEN mit: teXt   Wort   Klasse   Quelle

Fig. 8: Building a relationship between two terms in INDEX.

```
┌─────────────────────────────────────────────────────────┐
│   ┌───────────────────────────────────────────────────┐ │
│   │ KLASSIFIKATION          Nr:    5231  S:   M:      │ │
│   │ Notation              │ B T  Eb U │ 1 2 3 │ F     │ │
│   │ AfAa0005    ::FA      │   K    1  │       │       │ │
│   │ . * Aachener und Münchener Lebensversicherung AG  │ │
│   │ allgemein                                         │ │
│   │ ┌─────────────────────────────────────────────┐   │ │
│   │ │ WORT              Nr:    8428 H: 0 S:   M:  │   │ │
│   │ │ Aachener und Münchener Lebensversicherung AG│   │ │
│   │ │ ┌ Zusatz: ────────────────────────────────┐ │   │ │
│   │ │ │ VBKO                              < __ > │ │   │ │
│   │ └─────────────────────────────────────────────┘   │ │
│   └───────────────────────────────────────────────────┘ │
└─────────────────────────────────────────────────────────┘
```

Wort VERBINDEN mit: teXt   Wort   Klasse   Quelle

Fig. 9: Building a relationship between a classification and a term in INDEX.

```
┌─────────────────────────────────────────────────────────┐
│   ┌───────────────────────────────────────────────────┐ │
│   │ WORT                 Nr:     482 H: 0 S:   M:     │ │
│   │ Öffentliches Verkehrsmittel                       │ │
│   │ Sp │ Sgr Fa │ Wo G Gm Sch │ 1 2 3 │ Z1 Z2 │ F │ T │ │
│   │    │     G  │             │       │       │   │   │ │
│   │ ┌─────────────────────────────────────────────┐   │ │
│   │ │ QUELLE                   Nr:     5 S:x M:   │   │ │
│   │ │ Meyers Enzyklopädisches Lexikon in 25 Bänden│   │ │
│   │ │ Zitcode  Jahr DT E Sp  Geg   L Sgb 1 2 3 Standort│
│   │ │ MEYERS   1980 1x   de                    xyz│   │ │
│   │ └─────────────────────────────────────────────┘   │ │
│   └───────────────────────────────────────────────────┘ │
└─────────────────────────────────────────────────────────┘
```

Wort VERBINDEN mit: teXt   Wort   Klasse   Quelle

Fig. 10: Building a relation between a source and a term in INDEX.

relationship either from a list (by pressing F1) or by typing the number of the relationship into the clamp and pressing a function key.

At every stage the user is able to zoom (means: to display) the relations to a descriptor. He gets a list. The list he receives can be posed everywhere on the screen. It is possible to delete relations from the list, but not more than one relation at once.

It is also possible to connect terms to a classification card or to a source card by defining a special kind of relationship.

8. Limits and data structure

Regarding the limits of data input, there is a limit of 65.535 terms in TMS. The TMS data structure is something of a hybrid. Probably the nearest formal structure to which it could be compared is a relational model, however there are elements of the data which are purely hierarchically structured. The program becomes slower as more terms and relations are entered. In PROTERM and INDEX there are practically no restrictions regarding to the number of terms. PROTERM is designed with a developing tool called Clipper. The data structure is a simulated relational model. The program becomes much slower as one enters more terms and relations. INDEX is a purely semantic network. The access time to the data is very fast, it will not be reduced by the quantity of data entered.

9. Batch functions

The batch functions contain:
- Import and export of word material
- Import and export of structured material
- Output on different media
- Different kind of output

Before starting to construct a thesaurus it is more or less common to implement a list of words from other existing specific thesauri. These data exists normally in machine readable form, hopefully in ASCII format.

All three producers provide to convert the data into the special thesaurus software format. But big problems come up, if the data are already structured, because no standards for data exchange exists. ERNST LUKAS has set up a program to convert the already structured AGROVOC-Thesaurus of the FAO into INDEX format. Furthermore ERNST LUKAS SOFTWARE provides a data exchange program to convert data files from INDEX to the Data Bank Management System BASIS and they plan to do one for the system TRIP.

The author is not absolutely sure if PROGRIS provide similar exchange programs. Maybe it is possible with PROTERM-K (for Communication), a special modul.

10. Output functions

In all three programs the user gets highly developed output functions, especially in the very sophisticated INDEX and in PROTERM. In INDEX for instance the user can produce an output file of his thesaurus, ready for electronic publishing.

Apart from TMS all standardized lists are possible, for displaying on the screen, for saving as a file or for the output on the printer.

In addition, PROTERM and INDEX provides various sorting parameters.

4. Management and System Functions	INDEX Version 4.1	PROTERM-T Version 2.5	TMS Version 1.0
User Password and Restrictions	Yes	No	No
Display/Change of - Menus - Error Messages - System Messages - Colours	Yes Yes Yes Yes	No No No No	No No No No
Change of - Recording Order - Recording Labels/Display Labels	Yes Yes	Yes Yes	Yes No
Draw Up/Change/Display of Print Parameters	Yes	Yes	No
Reorganisation - Single Files - All Files	Yes Yes	No Yes	No Yes
Display of the Statistical Structure	Yes, Module	Yes, for all terms	Yes, per Term and Relation
Implemented Backup Routine	Yes	Yes	No
Management of more Thesauri	Yes	Yes	No
Management of a Multilingual Thesauri	Yes	No	Yes, up to 2 additional L.
Change to Operation System Level without leaving the Program	Yes	No	No
Change to a Word Processing system without leaving the Program	Yes	No	No

Table 4: Management and system functions

The different output lists are:
(I=INDEX; P=PROTERM; T=TMS)

– Alphabetic term list	I	P	T
– Systematic term list	I	P	
– Minithesauri/Partthesauri	I	P	
– Termlists with/without relationships	I	P	T
– Termlists selected by different criteria	I		
– Synonymlists	I	P	
– Systematic thesaurus	I	P	T
– Alphabetic thesaurus	I	P	T
– Thesaurus with different BT/NT levels			T
– Hierarchical BT/NT lists		P	
– different classification lists	I		
– different source register lists	I		

11. Management and system functions

All three programs provide various management and system functions, as shown in Table 4.

12. Conclusion

In this article the author attempted to present some results of his evaluation of the functions of three thesaurus software products. Each program complies more or less with the minimum standards of a thesaurus software. If ergonomical factors of software and additional features are taken into consideration the INDEX program turns out to be more efficient than PROTERM and TMS.

PROGRIS with the demand to serve their users with an acceptable product for less money is on the right line. Regarding the relation between price and performance they produce a good software.

How does higher cost compare with high performance? That is not always the case. TMS is much more expensive than PROTERM but its performance is only half of PROTERM.

Naturally there is a third factor: the suitability of the software. All three programs are reasonably good for the establishment of small thesauri, but if one needs a powerful multiuser system for the management of extensive archive material or for constructing a broad multilingual or even a monolingual thesaurus one should take PROTERM rather than TMS. In this case INDEX with its highly sophisticated features would be the preferable alternative.

References

(1) Ritzler, C.: Vergleichende Untersuchung von PC-Thesaurus-programmen. Diplomarbeit. Fachhochschule Darmstadt, Fachbereich Information und Dokumentation. Darmstadt 1989. 156 ref. 200 p.

Dipl.Inform.wirt Claus Ritzler,
Lünkenweg 6, D-2000 Hamburg 60

Klassifikation, Datenanalyse und Informationsverarbeitung

The 15th Annual Conference of the Gesellschaft für Klassifikation is planned from Febr. 25-27, 1991 at Salzburg, Austria with the topic *Classification, Data Analysis and Information Processing*. It is understood as being a challenge for interdisciplinary research. Relevant application fields have been pinpointed, such as archeology, biology, medicine, linguistics, geography, law, technology, economics, social science on the conception of knowledge and databanks, expert systems and data analysis packages, as well as terminology, documentation, subject analysis and library classification. 12 areas with 48 subtopics are outlined in the Announcement which calls for papers until Oct.30, 1990 to the Conference Chair, Prof.Dr.H. Goebl, Institut für Romanistik, Universität Salzburg, Akademiestr. 24, A-5020 Salzburg, Austria.

3rd IFCS Conference, Edinburgh

A second announcement and Call for Papers has been released for the Third Conference of the International Federation of Classification Societies to be held from Aug.6-9, 1991 at the Heriot-Watt University, Edinburgh, Scotland. Contributions from all areas of statistics are expected, theoretical and applied subjects will be covered. Invited speakers will include E. Diday (France), R.C. Dubes and D. Pregibon (USA), B.S. Everitt and P.H.A. Sneath (U.K.), and N. Ohsumi (Japan). Other invited Sessions will cover:

Tree theory - Cluster analysis - Combinatorial optimisation - Diagnostic keys - Phylogeny - Classification in psychology - Geological classification - Multivariate longitudinal data - Sequencing - Classification of living organisms - Computing methods.

Abstracts in 3 copies are due by Jan.15, 1991. They should be sent to Prof.David J. Hand, Faculty of Mathematics, the Open University, Walton Hall, Milton Keynes MK7 6AA, England.