procedures to be used by the system. A large number of additional parameters controlling how the INTERBAT system operates were also defined during implementation but these need not concern most users of the system.

BASIS was chosen from a variety of commercially available text information retrieval packages mainly because it was capable of providing all the search features required for INTERBAT, it was available on a bureau which could be accessed from the other research laboratories of the BAT Group and it was directly compatible with the "in-house" computer which will eventially be available within GR&DC.

It will be possible to handle extensions to INTERBAT and indeed any other text information requirements by using further implementations of the BASIS software package. These further implementations would involve the setting up of additional computer files of information and their associated indexes, but the basic search and update facilities provided by the BASIS package would be the same as for the trial INTERBAT system. The main search procedures in INTERBAT are described briefly in Section 7. However a BASIS users manual is available and describes in detail all the search and update facilities provided by the package.

4. THE STRUCTURE OF REPORT RECORDS

The ~~pilot~~ trial INTERBAT system is specifically designed for bibliographic data about R&D reports generated within the BAT Group. It is intended that the system will be extended to hold other types of information after the trial system has been assessed.

For each report the trial INTERBAT system holds a record or document containing the following fields of information.

Get A.no.

~~Delete AV, TI, RN~~

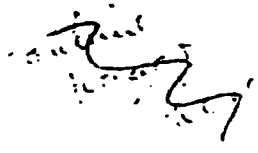
~~or ** 2, ** 4, ** RN~~ 6.

AT ** 2

Field Number	Contents	Short Field Name
1	Accession number	AN
2	Authors	AU
3	Source organisation	SO
4	Title	TI
5	Publication date	PD
6	Report number	RN
7	Source language	SL
8	Translations available	TR
9	Index terms or descriptors	IT
10	Abstract or summary	AB
~~11~~	~~Security code~~	~~SC~~
12	HOUSEKEEPING	HK

YES.

The short field names are used to refer to the fields of data when searching for information using INTERBAT. Details of the way in which information for each field is coded are given in Section 5. A sample record from the database is shown in Figure 1.



three additional short names may be used to refer to commonly used groups of fields (called MAPS) when displaying retrieved information. These are S1 referring to author, title and report number, S2 referring to the S1 fields and the abstract and S3 referring to title, index terms and abstract.

1014677781

The records constitute the basic information stored in the system but, additionally, indexes are also kept for each field except fields 7, 8 and 11 (the source language, translations and security code). Thus each unique accession number, author, source organisation, publication date report number and descriptor is indexed automatically with pointers to each record in which they appear. For the title and the abstract each individual word is indexed except for a "stopword" list of common words such as BUT, AND, THE, HOWEVER etc. The production of these indexes ensures that retrievals may be executed rapidly without searching through the information records themselves.

The "stopword" list will be extended from time to time and the whole database reindexed. In the meantime some trivial words will inevitably appear in the word indexes for the title and the abstract.

5. THE CODING OF INFORMATION IN THE REPORT RECORD

The previous section defined the way in which the report record is composed of fields of information. This section describes the way in which the information in each field is coded or produced.

ACCESSION NUMBER (AN)

This number is automatically assigned by the system as each new report record is entered onto the system. The accession numbers are assigned sequentially and if a record is subsequently deleted its accession number ceases to be used.

AUTHORS (AU)

All authors names are recorded. The surname appears first, then a comma followed by the initials separated and terminated by fullstops. A semicolon is used to separate authors, e.g. SMITH,A.B.;JONES,C.D.

SOURCE ORGANISATION (SO)

This field contains a code representing the organisation (laboratory) at which the report was produced. Codes in use so far are as follows:

Code	Source Organisation
AMAT	R&D Dept., AMATIL Ltd., Sydney, Australia.
BWTC	R&D Dept., B. & W. Tobacco Corporation, Louisville, U.S.A.
GRDC	GR&DC, Southampton.
HBAT	R&D Dept., BAT, Hamburg.
ITLM	R&D Dept., ITL, Montreal, Canada

TITLE (TI)

The title field contains the title as it appears on the front of the report.

PUBLICATION DATE (PD)

Dates are recorded as six digit integers consisting of the last two digits of the year, a 2 digit month and a 2 digit day. For example 10th March ~~10~~ 1980 is recorded as 800310.

REPORT NUMBER (RN)

These are generally recorded as shown on the report title page. If a word appears after the number it is replaced by a hyphen and the first letter of the word. For example RD.1685 Restricted is recorded as RD.1685-R.

Hamburg progress report numbers are coded as on the report except that they are preceded by HPR-. For example, the 1979 April to June progress report is recorded as HPR-II/79.

SOURCE LANGUAGE (SL)

This field contains a two character code for the language in which the report was originally written. The codes used are those used in the

LOOK AU = A**.

LOOK AU = A*

BIOSIS file of the SDC/ORBIT system. See Table 1 for a list of languages and source codes.

TABLE 1

LANGUAGE CODES USED IN THE SOURCE LANGUAGE AND TRANSLATION FIELDS

Name	Code	Name	Code
Afrikaans	AF	Kirghiz	KI
Albanian	AB	Korean	KO
Amharic	AM	Loatian	LO
Arabic	AB	Latin	KZ
Armenian	AR	Latvian	LA
Assamese	AS	Lithuanian	LI
Austrian	(See German)	Macedonian	LU
Azerbaijani	AZ	Malay	MA
Basque	BA	Malayalam	ML
Belorussian	BE	Malayo-Polynesian	MP
Bengali	BN	Marathi	MM
Bohemian	(See Czech)	Mexican	(See Spanish)
Bulgarian	BU	Moldavian	MO
Burmese	BV	Mongolian	MG
Cambodian	CA	Nepali	NP
Catalan	CT	Netherlandish	NE
Chinese	CH	Norwegian	NO
Croatian	CR	Panjabi	PA
Czech	CZ	Persian	(See Farsi)
Danish	DA	Polish	PO
Dutch	(See Netherlandish)	Portuguese	PT
English	EN	Provencal	PR
Esperanto	EP	Pushto	PU
Estonian	ES	Romanian	RO
Faeroesse	FA	Russian	RS
Farsi	PE	Sanskrit	SA
Fijian	FG	Scottish	(See Gaelic-Scottish)
Finnish	FI	Serbian	SE
Flemish	(See Netherlandish)	Serbo-Croatian	SR
French	FR	Sindhi	SI
Gaelic (Irish)	(See Irish)	Sinhalese	SH
Gaelic (Scottish)	GA	Slovak	SL
Georgian	GK	Slovenian	SN
German	GE	Spanish	SP
Greek	GK	Swahili	SW
Gujarati	GU	Swedish	SS
Hebrew	HE	Tadzhik	TA
Hindi	HI	Tagalog	TG
Hungarian	HU	Tamil	TM
Icelandic	IC	Thai	TH
Indonesian	ID	Tibetan	TI
Interlingua	IG	Turkish	TK
Irish	IR	Turkmen	TU
Italian	IT	Ukrainian	UK
Israeli	(See Hebrew)	Urdu	UU
Japanese	JA	Uzbek	UZ
Javanese	JV	Vietnamese	VI
Kannada	JZ	Welsh	WE
Kazakh	KA	Yiddish	YI

TRANSLATIONS AVAILABLE (TR)

This field contains language codes for languages into which the report has been translated. The relevant codes are separated by semicolons if more than one language translation is available. The codes used are as for the source language field.

INDEX TERMS OR DESCRIPTORS (IT)

Index terms (descriptors) will be recorded in this field and separated by semicolons. A procedure for assigning descriptors has not yet been agreed and the field is currently left blank.

ABSTRACT OR SUMMARY (AB)

The abstract or summary of the report is recorded on the system essentially as it appears in the report. If tables are given in the abstract or summary of the report an attempt has been made when coding to convey the tabular information without using a tabular structure since this would be distorted by the variable length output formats.

SECURITY CODE (SC)

It is possible to assign different security levels to records of information and link these to user's security keys so that different users may be given access to different information. At present access is not being limited in this way, but the facility may be invoked if required at a later date.

6. DATA ENTRY AND COVERAGE OF INFORMATION

New Reports

For the initial stage of the trial INTERBAT system the following procedure has been adopted. The title and summary pages of all non-confidential R&D reports issued in GR&DC are sent to the computing section

at the time of issue, for entry into the system. Similar data is sent for other R&D reports as they are received in the GR&DC library. The relevant data is ~~then~~ punched onto computer cards and read into the INTERBAT system using the field terminal in GR&DC.

~~The majority of reports~~ A total of 132 reports issued or received in GR&DC since the beginning of 1979 have now been put onto the INTERBAT system in this way.

Pre-1979 Reports

It would have been an almost impossible task to code and enter the information for the complete backlog of research reports issued within the BAT Group before 1979. However, B. & W. have been operating a local system for many years so information on a large number of reports already existed in computer readable form. B. & W. agreed to provide the non-confidential information on a magnetic tape and a special program was prepared to transcribe the pre-1979 data from the B. & W. tape into a form suitable for entry into the trial INTERBAT system.

The B. & W. data on each report consisted of a catalog field and a text field. The catalog field contained a report number which was similar to the one appearing on the original document but with a two digit year indicator inserted after any leading alphabetic characters in the report number (for example RD.1251 is recorded as RD75-1251). The text field contained an abstract, the title, the authors names and any associated keywords. Since this text information was not divided in any systematic way it was not possible to separate automatically authors or titles for transciption into the appropriate fields in the new system and a manual separation would have been sufficiently time consuming to defeat the object of obtaining the B. & W. tape.

The program to transcribe the B. & W. tape therefore adopted the following procedure.

The report number in the catalog field of the B. & W. record was examined and its format used to ascertain the source organisation and hence the source language. The report number also gave the year of publication so that a crude publication date of January 1st for that year was generated and placed in the publication date field of the INTERBAT record. The report number was placed unchanged in the report number field of the INTERBAT record.

Since no automatically discernible title was available the report number was repeated in the title field so that a searcher would obtain at least some useful information if listing titles from a group of retrievals records.

The whole of the text field on the B. & W. system was copied to the abstract field on the INTERBAT system and the words SEE ABSTRACT were placed in the author field.

An example of a resulting INTERBAT record generated from a B. & W. tape record is shown in Figure 2. Data for a total of 3000 pre-1979 reports were available on the tape and these have now been transcribed into the INTERBAT system. Of the 3700 reports, 2000 were GR&DC RD reports, 500 were B. & W. reports and 100 were Hamburg reports.