PRICES AND ONLINE SEARCHING: THE CHEAPEST HOST IS NOT ALWAYS THE ONE YOU THINK

T. Baiget, CIDC, Spain

Keywords: Online searching, Costs, Telephonic prices, hosts, System quality, Criteria, Language flexibility. Abstract: When the same file is available from several hosts, choosing the host offering the best quality-price ratio is not a straightforward operation. Often the announced file prices are just the top of the iceberg. Quantity of data available, eventual extra costs regarding online and offline prints, administrative constraints, royalties are to be looked at carefully. Also the computer response time, search techniques, telecomunications costs, the access speed, etc. are decisive on the final bill. A detailed analysis is shown. Some results are given for the most important bibliographic files currently available from the main hosts. The detailed explanation will help the online users at present and in the future in choosing their hosts according to their needs and their geographical location.

1 INTRODUCTION

In face of the great number of available hosts that offer the same or similar databases, the user must choose an option.

Till now some studies have been carried out to that purpose but most of them only compare American hosts- Bibliographic Retrieval Service (BRS), DIALOG and System Development Corporation (SDC/ORBIT).

In Europe there are very interesting and competitive hosts whose quality/cost relationship is sometimes better than that of the the Americans.

The European user sometimes intuitively knows the advantages of these hosts but often she or he is reluctant to assume the risk, or simply has no time to <u>invest</u> in an European choice.

It would be extremely complicated to establish a ranking of all available hosts in terms of cost and quality, since the number of parameters involved is countless, and beyond a certain point its importance begins inevitably to have a subjective component.

Nevertheless, at first sight some of these parameters seem to be clearly more important that others, as for example telecommunication costs, processing time, etc.

The Consorci d'Informació i Documentació de Catalunya (CIDC)

has some years' experience in connecting to more than 15 hosts. The CIDC is a multidisciplinary intermediary centre, with tariffs partially funded by the local government of Barcelona, open to all users. These come in the same proportion from University, Administration and Industry. This situation of free demand has spontaneously shaped a particular kind of usage of the various European and American hosts producing a "natural selection" among them.

There are evident limitations in this work and we apologise for the probably numerous lacks it has. Its aim is not an almost-impossible multicomparison of hosts but an exposition of points of view that might help to evaluate the costs. Although several hosts are considered, the comparison is established mainly between the two big competitors in the European market: DIALOG and ESA/IRS.

We believe that this paper may at least be useful to point out some of the main parameters to bear in mind when choosing a host, as well as an invitation for the user to consider if she/he is really using the best option available.

2 NOMINAL PRICES OF DATABASES

The first approach to compare online search costs of different vendors is to compare their access fees. Since connection to the host is always carried out through a telephone line, the cost of this has to be taken into account together.

Let us see some examples of the international telecommunication costs involved in Europe to have access to different hosts at 300 baud.

Country	To Europe		To USA	
		US\$/h		US\$⁄h
*France	0,4FF/min+l2FF/Kseg	4.23	lFF/min+30FF/Kseg	10.54
*Germany	0.llDM/min+5DM/Kseg	3.36	0,30DM/min+16DM/Kseg	9.48
Spain	12pta/min+380pta/Kseg	7.68	1000pta/h+50pta/Kc	27.40
*Sweden	lSKr/min+20SKr/Kseg	11.01	2SKr/min+60SKr/Kseg	23.06
*U.K.	2,2p/min+1,2\$/Kseg	2.98	8p/min+3£/Kseg	10.08

Kc= 1000 characters
Kseg= 64000 characters

* Monthly fixed fees have not been taken into account.

TABLE 1. PTT prices (bilateral, EURONET, TYMNET, TELENET).

On the right side of the TABLE 1, binomial tariffs have been converted to the average cost of one hour in US dollars taking into account the characteristics of the online searching type of connection **.

** It has been estimated a typical average search lasting 15 minutes, consisting of 25 descriptors, 5 online prints of titles and descriptors, 5 online prints of only titles and a saved strategy executed once. This means the transmission of c.a. 5000 characters in any direction, i.e. 20.000 characters per hour. Nevertheless, the best way to know the real cost per minute is to check the invoice received from PTT, because sometimes there are surprises.

For instance, in Spain, an expected cost of 35 pta/min for American calls rises to approximately 45 pta/min because the Spanish PTT rounds the usage to periods of 6 minutes and blocks of 1000 characters.

Minimum charges per call are frequently applied by PTT. Characters transported are usually measured in segments. One segment equals 64 characters, i.e. half a packet of 128 characters. In this cases, the minimum charges are based on minutes and blocks of 10 segments (640 char.).

When calculating telephone costs based on segments one must know that the PTT charge is the same for transmitting (or receiving) one character or for any quantity up to 64. If one sends 65 characters, one pays as much for 128. Therefore, long search statements are cheaper than short ones. This is an advantage for the retrieval languages that allow to stack commands as ESA/QUEST, DIALOG and ORBIT, and to certain extent GRIPS/DIRS-3.

Rolf Jansen (Ref.10) states that in Germany the segment tarification leads to an increase of 60% in the volume factor of the binomial tariff. When receiving their PTT invoices users find that the real cost of their calls to USA, for example, has not been 1.6 pf/segment but 2,56 pf/segment.

ESANET

Country	TO FRASCATI (ESA/IRS)
Denmark France Italy	<pre>} 10 AU/h≃9.8 US\$/h</pre>
U.K.	(1 AU ≃ 0.978 US\$, 1983)

DIALNET

Country	TO PALO ALTO (DIALOG)	
U.K.	15 US\$/h	

TABLE 2. Special links.

Leaving apart some special links (TABLE 2), and considering the PTT figures given in TABLE 1, one can state that on the average, telecommunications to European hosts cost approximately one third of those to American ones.

For the calculations in TABLE 3, an average of the PTT prices has been done. For international European connections we have taken the value 5.85 US\$/h, and for American ones 16.11 US\$/h.

TABLE 3 shows access + telecommunications costs of one American host- DIALOG - and three European - ESA/IRS, DATA STAR and DIMDI, for some core data bases.

According to various studies (Refs. 5, 6, 9) among BRS, DIALOG and SDC, BRS is the cheapest system in the USA, whereas DIALOG has the best monthly discounts scheme for heavy usage.

Europen nominal access fees are in general lower than the American. If moreover telecommunication prices are considered one can conclude at first sight that, at least for the European users, the European hosts are more competitive.

There are some important aspects to take into account from TABLE 3:

- Data base time span.

It is not the same thing a COMPENDEX on DATA STAR beginning on 1976 or INKA (1975), as on SDC and DIALOG (1970) or ESA/IRS (1969).

- Data Base fragmentation.

Splitting data bases in various parts is something that can affect seriously the search costs, as the user is compelled to repeat the search strategy in every part. There are three cases of segmented data bases:

- 1. All parts permanently online, as it is the case of SDC, DIALOG, BLAISE, INKA, etc.
- 2. Parts available under a time schedule, f.i. DIMDI (BIOSIS 70-75, etc.), TELESYSTEMES (FRANCIS).
- 3. Some parts permanently offline, f.i. DIMDI (ISI/BIOMED, etc.), NLM.

In this last case, users carry out the search on the online segment and then they order the search to be continued offline on the other segments. The problem of this is that they do not know the results until receipt of prints.

The case of ESA/IRS, whose databases have no fragmentations has to be urderlined. CHEMICAL ABSTRACTS (5,6 million references) and PASCAL (4,2 million references) are the world's biggest unique bibliographic files.

- Online prints.

An increasing number of database producers also charge a royalty for online prints. Nevertheless most hosts have trial formats to check the intermediate search results for which there is no royalty.

- Offline prints.

In some hosts the announced print costs are not inclusive of postal charges. Taking into account that prints are normally sent by express airmail, postal costs may often represent 20% of the print costs, depending on the databases (DIMDI, DATA STAR, etc.).

Minimum fees, Subscriptions and Joining fees.

Some hosts charge a quantity regardless of whether you use the system or not (DIMDI=DM100/quarter, NLM= £ 10/month, SDC=10 US\$/month if used that month). Finally, some hosts, mainly those specialised

TABLE 3. Some significant nominal cost differences among some hosts that duplicate core files. Average telecommunication costs have been added. Hourly rates and prints in US dollars. * An A4 page = 4 references without abstract or 2 refs. with abs. An A5 page = 2 refs. without abs. or 1 ref with abs. 20% for postal charges has been added.

ESA/IRS			DATA ST.	AR		DIMDI			DIALOG		
File	Access+ Telecom	Print off	File	Access+ Telecom	Print off *	File	Access+ Telecom	Print off *	File	Access+ Telecom	Print off
2 CHEMABS 67-	65	0.20	СН08 67-71 СН09 72-76 СНЕМ 77-	60	0.18				308CASEARCH67-71 309 " 72-76 320 " 77-79 310 " 80-81 311 " 82-	80	0.20
7 BIOSIS 69- No abstr.	58	0.18	BIOSIS 70-77 " 78- No abstr.	63	0.06	BIOSIS 70-75 " 76-77 " 78- With abstr.76-	47	0.14	55 BIOSIS 69-76 5 " 77- With abstr.76-	74	0.15
4 COMPENDEX 69-	75	0.25	COMPENDEX 76-	78	0.34				8 COMPENDEX70-	96	0.30
8 INSPEC 71-	68	0.23	INSPEC 70-77 " 78-80 " 81-	73	0.31				12 INSPEC 69-76 13 " 77-	91	0.25
						EM BASE 74-77 " 78-	57	0.18	172 EXC.MED74-79 72 " 80-	86	0.20
			MEDLINE 66-70 "71-74 "75-78 "79-	32 (43 in 1983?)	0.10	MEDLARS-1 64-65 MEDLARS-2 66-75 " 76-77 " 78-	26	0.05	152 MEDLINE66-72 153 " 73-79 154 " 80-	51	0.15
16 CAB 72-	66	0.17		-		CAB PLANTS 73- CAB ANIMALS 72-	54	0.18	50 CAB 72-	66	0.30
6 NTIS 64-	52	0.10	NTIS 70-74 " 75-80 " 81-	52	0.13				6 NTIS 64-	56	0.10

in business, charge a joining fee that can be important. - Restrictions.

There are databases whose access is restricted to some users: DOE/EDB, TRADE OPPORTUNITIES, TRIS (DIALOG) restricted to Europe; VETDOC, RINGDOC, PESTDOC, CRDS, TULSA (SDC) restricted to printed version subscribers, etc.

But nominal prices are only one factor to consider. The other factor is the time needed to be connected to the host to get a complete and satisfactory answer. There are hosts that having cheaper nominal prices result much more expensive due to their poor retrieval language or their slow processing time (Ref. 11).

3 PROCESSING TIME

Up till now we have been studying the visible part of the iceberg, and also some aspects that sometimes are covered by the waves, i.e. they are not clearly explained in the brochures and general information. Let us go deeper studying the system performance.

The comparison of the quality of the different hosts is very difficult because of the large number of parameters involved.

But it is possible to simplify the problem, and to introduce a new approach to the objective of determining real costs, that is the measurement of the response time to a given strategy.

The easiest hosts to compare are ESA/IRS and DIALOG, because their retrieval languages are very similar since both come from the old RECON.

In order to avoid the typing time, a search strategy has been pre-recorded in an intelligent terminal (an Apple II with a Visiterm programme).

The strategy (about "Hazard of the static electricity") was:

S	HAZARD?	
S	RISK?	
S	DANGER?	
S	SAFE?	
C	1-4/+	
S	SPARK?	
S	ELECTROSTATIC?	
S	STATIC (W) ELECTR	ICITY

SHAZARD?/TI,DE,ID SRISK?/TI,DE,ID SDANGER?/TI,DE,ID SSAFE?/TI,DE,ID cl-4/+ SSPARK?/TI,DE,ID S ELECTROSTATIC?/TI,DE,ID SSTATIC(W)ELECTRICITY/TI,DE, ID SSTATIC(W)CHARGE?/TI,DE,ID C6-9/+ C5*10

SSTATIC(W)CHARGE? C6-9/+ C5*10

conducted on COMPENDEX. Result: 95 references on ESA, 93 refs. on DIALOG. September 1982.

Tests have been carried out in the morning during one month, and times obtained from 30 measures on each system are the following:

ESA/IRS	DIALOG
0.75 sec	2.71 sec

DIALOG answer time is still worse in the European afternoon when in America it is the morning.

A problem observed on DIALOG when sending pre-recorded strategies of this size (that of the example has 181 characters) is a delay of approximately 1 minute in giving the answer to the first command, whatever way is used to reach the computer. Because of this, when searching on succesive databases it is quicker to end/savetemp the strategy than to send again the pre-recorded strategy from the terminal.

If instead of COMPENDEX the comparison is established with CHEMICAL ABSTRACTS, the difference in time and in cost is dramatic:

If the example search is made in ESA/IRS, one gets the 208 references on "Electrostatic hazards" in 1 minute, while if it is carried out in DIALOG for the same period (1967-1982) one spends 11 minutes (searching in one segment, end/saving, and executing the strategy in the other four segments). As DIALOG is more expensive, the total differences in cost are still higher.

In order to avoid the delay that the reception of prerecorded strings cause in the DIALOG system, there have been accounted the processing time of EXECUTING STEPS the same strategy recorded with the END/SAVE command. In this case, the net processing time has been:

DIALOG						
1.88	sec					

The difference 2.71 - 1.88 gives the 50 seconds of delay produced when using an intelligent terminal.

According to this results we can say that ESA/IRS is 2.5 times faster than DIALOG. Of course this is only net processing time. In the real searches other times are involved (thinking, typing refinements, etc.) and the total difference is not so great, as we shall see.

4 OTHER COST-EFFECTIVENESS CONSIDERATIONS

As it has been said, the evaluation of the various languages and features available from different hosts is very complicated and frequently its value depends on the kind of information to be retrieved, the search methods, the particular preferences of the searcher, etc.

In our opinion there are in general two main characteristics that determine the attractiveness of hosts to the users:

a) Quantity of data available.

b) Retrieval language flexibility.

a)The number of databases and references measures the desirable self-sufficiency of a system to answer a demand with adequate recall. Searchers do not like the need to have access to two or more hosts to complete the same demand, having to re-prepare and re-type the strategy, and increasing costs.

If the databases chosen are in the same host it is possible to save the strategy and to pass it automatically on other databases.

1	From	the	poin	t of	view	of	data	available	(Sept.	82)	hosts	can
be	arra	anged	l as	shown	ı in	TABL	E 4.					

Host	N°of databases	N°of million refs.
DIALOG	134	59
SDC	69	35
ESA/IRS	31	22.1
DIMDI	23	20.2
DATA STAR	16	18.3
TELESYSTEMES	29	8.4
INKA	23	6.5
BLAISE	7	2.5

TABLE 4. Host ranking according to their quantity of data available. Splitted parts, training and parking databases have not been considered.

TABLE 4 shows the predominant situation of DIALOG that, as a big information hypermarket, is the unique host that is nearly self-sufficient in a multidisciplinary basis.

When a large number of databases is not possible, the specialisation of hosts is a good approach to self-sufficiency. A host with all its files related to the same field of science is much more attractive than one with various little samples on different fields.

In Europe there are two main specialised hosts that are selfsufficient in most of the searches on their broad fields: DIMDI on life sciences, and ESA/IRS in science and technology.

Chemistry is a particular field in which a satisfactory answer is attained usually with only one file: Chemical Abstracts.

DIALOG	DIALOC	U	Ŋ				
SDC	29	SD	/IR		~		
ESA/IRS	21	9	ESA	IQ	STAF	EMES	
DIMDI	11	4	3	DIN	TA	TST	
DATA STAR	11	8	6	3	DA	ILES	
TELESYSTEMES	-	-	2	-	-	ΤĒ	ХA
INKA	5	3	4	-	2	-	INI
BLAISE	-	1	-	-	-	-	-

TABLE 5. Number of databases duplicated among some hosts.

It could be thought that the usage of a host is proportional to its number of databases or references. Some factors in the practice cause the attractiveness of the big hosts to be higher than a proportional rate.

One is economic, based on the discounts offered to the users when contracting a guaranteed minimum number of hours, and on the rebate schemes applied that offer more discounts for larger number of connected hours.

But the most important factor lies in the searcher habits. If one host has more databases the user must have access to it more often thus gaining more confidence and skill with it.

Other hosts with different information may be interesting but they have to be attractive enough to counteract the user reluctance to change to a different language, studying a thick manual and feeling unconfortable during the search.

Nevertheless, although this reluctance is decisive for the usual daily work, there are occasions that the connection to another host is mandatory in order to solve a search satisfactorily. Then, most information centres have to have access to various hosts, using one or two for the heavy usage and two or three more for the marginal searches.

An interesting phenomenon has happened in Europe with hosts with the same retrieval language. The maximum competitors DIALOG and ESA/IRS have become really allied in the end, thanks to their similarity that permits users to change easily from one to the other, and thus taking profit of the advantages of both. ESA/IRS had prepared the field when DIALOG arrived to Europe in 1974, and at present ESA/IRS, to some extent, benefits of the essentiality of DIALOG.

A similar case has happen, to a lower level, between SDC and BLAISE.

b) Online searching is an art where the skill and good shot of searcher are fundamental. The interactiveness of online systems aims at the correction of bad shots, approaching the results to the intended target. But not all the retrieval languages are equally able to allow the introduction of modifications in the search strategy.

There are three kinds of potential refinements or modifications to a performed strategy:

1.To adjust the subject when inadequate keywords have been used. All systems allow one to enter new terms, ignoring or deleting those entered previously and building a new strategy. 2.To improve the precision. The availability or not of distance operators establishes a clear division among hosts, measuring their quality. This feature is not available on TELESYSTEMES and BLAISE yet, and it is currently being implemented in successive files of SDC.

3.To adjust the quantity, (and also the precision) specially when a too high number of hits has been obtained. Sometimes this operation takes more time than the whole search.

To obtain the desired/expected point in the recall-precision curve, the searcher does his/her guesses and normally, in any system, is able to choose a more or less restrictive strategy on more or less restrictive fields ranging from titles to abstracts.

There is a variety of situations among hosts about the inverted fields, and frequently there are differences between databases of the same host.

The benefit of the abstract inversion is highly questionable. In our opinion the number of searches in which it is necessary to search in the abstract field is negligible. In any case, abstracts can be inverted only if the system allows the possibility to eliminate them from the search at will. Otherwise, the search is inadmissibly imprecise.

Some systems allow search limitation to some specified fields at once at the beginning of the search (SDC, ESA/IRS, etc.). , but in others, fields must be indicated tediously in every term entry (DIALOG, etc.). If in this case the precision is not correct, the user is compelled to retype the strategy.

Finally there are systems that have the greater flexibility, allowing users to restrict the search results at the end of the search, after they know the number of retrieved items (ESA/IRS). This feature is enormously important and efficient since it eliminates the searcher's need to guess which is the more adequate level of precision: titles-descriptors-identifiersabstracts, and therefore eliminates the risk of having to retype the strategy.

Many other considerations could be made to compare host efficiency but finally the solution has to be found in a statistical analysis of the daily practice.

An statistical analysis has been carried out recently at CIDC (11). The study includes data from several hosts but the only statistically valid samples are the data of ESA/IRS and DIALOG.

	n ^o of entries	Mean time per entry
ESA/IRS	279	6.23 minutes
DIALOG	380	8.34 "

Over a period of 6 months the results show:

TABLE 6. Mean time per database used on ESA/IRS and DIALOG. Parking files have been ignored. End/saved strategies have been considered as entries except in split files.

REFERENCES

(1) Optimizing search costs. A comparative study of three systems to find the best one ... or two ... or three. (BRS, DIALOG, WISE). Folke, Carolyn Online, Apr. 81, p 38-43

(2) Online bibliographic services: a comparison (BRS,DIALOG, DROL, SDC). Weiss, S. Special Libraries, v 72, n4, p. 379-389, 1981. (3) Evaluation of Psychological Abstracts using three online systems: comparison of the results of a literature search. (BRS, DIALOG, SDC). Page, Alain Documentation et Bibliotheques, v26, n2, Jun 80, p 83-95. (4) Searching the chemical literature via three on-line vendors: A comparison (BRS, DIALOG, SDC) Ross, JC Journal American Society Information Science, v30, n2, p 103-106. 1979. (5) A comparison of three commercial online vendors. (BRS, DIALOG, SDC). Hoover, Ryan E. Online, v3, n1, Jan 79, p 12-21 (6) Comparative study of bibliographic online information systems. (BRS, CAN/OLE, DIALOG, SDC). Perusse, Lyse. Sainte Foy, Quebec University, 1979, 191p. Eric Report ED184578 (7) Experimental comparison of boolean operators in bibliographic retrieval systems. (DIALOG, SDC). Areu, Eloy; Brewer, Christina; Shneiderman, Ben. Proceedings of the ASIS Annual meeting, 1978. Volume 15. 41 st. Annual Meeting, New York, November 13-17, 1978, p13-16. (8) Some differences between three online systems: impact on search results. (BRS, DIALOG, SDC). Rouse, Sandra H.; Lannom, Laurence W. On-line Review, vl, n2, Jun 1977, p 117-132. (9) The new prices. Some comparisons. (BRS, DIALOG, SDC). Bement, James H. Online, vl, n2, Apr. 1977, p 9-22 (10) Wünsche der Online-Benutzer on die Deutsche Bundespost. Jansen, Rolf Vortrag auf der 4. Fruhjahrstagung der Online-Benutzergruppe der Deutschen Gesellschaft fur Dokumentation. 26 May 1982. (11) Modelo de control estadístico y administrativo de las búsquedas bibliográficas del Servicio de Teledocumentación del CIDC. Análisis de resultados. Melendez, Vicenç Reunión de Especialistas de Teledocumentación. Madrid. October 1982. (12) Some factors influencing choice of host systems. Batten, A.M. Proceedings of the 5th Online Information Meeting, London, 8-10 Dec. 1981. p 33-38