Decisive Factors in Choosing a Host System for Online Searching*

von Alexander *Nevyjel*, Seibersdorf und Otto *Oberhauser*, Wien

Entscheidungshilfe;
Host;
Information Retrieval;
Kosten;
Informationsvermittler;
Psychologie

Zusammenfassung

Entscheidungskriterien für die Auswahl eines Host-Systems zur Online-Recherche

Immer mehr Datenbasen werden auf immer mehr Host-Systemen verfügbar; dieser Markt wächst sowohl auf der Angebots- als auch auf der Nachfrageseite. Eine Informationsvermittlungsstelle hat üblicherweise Zugriff zu einer Reihe verschiedener Anbieter. Wenn daher ein Rechercheur die passenden Datenbasen für eine konkrete Online-Suche ausgewählt hat, betrifft seine nächste Entscheidung die Auswahl des Hosts, bei dem recherchiert werden soll. Das naheliegendste Entscheidungskriterium „Kosten" ist oft nicht das endgültige, da die tatsächlichen Kosten aus einer Reihe von Gründen schwer zu berechnen sind. Darüber hinaus weisen die globalen Jahresstatistiken der Autoren darauf hin, daß die durchschnittlichen Recherchekosten zwischen den einzelnen Hosts nur mäßig variieren. Auf der anderen Seite ergeben Fragebogen-Analysen bei Endbenutzern große Qualitätsunterschiede von Suche zu Suche, sofern „precision" und „recall" als Maße dafür gelten können. Dies muß daher in Betracht gezogen werden, falls die Auswahl eines Hosts einen Einfluß auf „precision" und „recall" besitzt. Auch weitere Faktoren wie Aktualität und Art der Aufbereitung einer Datenbasis, regionale/lokale Erwägungen, Postwegzeiten der Offline-Ausdrucke, Verfügbarkeit und Verläßlichkeit des Host-Systems sowie die speziellen Eigenschaften der Abfragesprache können für die Wahl des Hosts von Bedeutung sein. Die tägliche Praxis von Informationsvermittlern bestätigt jedoch, daß zusätzlich zu rationalen und finanziellen Aspekten auch psychologische Faktoren die individuelle Entscheidung zugunsten eines konkreten Systems beeinflussen.

Summary

More and more databases are available on more and more host systems. The market is growing in terms of supply as well as of demand. Usually, an intermediary service has access to a range of various vendors. So, if the searcher has selected specific databases fitting to a given query, his very next decision is the choice of the host. The most obvious decisive factor "cost" is often not final since real cost is difficult to calculate, depending on response time, comfort of query language, splitting of databases, form of contract, actual exchange rates, etc. Moreover, the authors' global annual statistics of cost per online session show only moderate variations between vendor systems. On the other hand, if precision and recall are accepted as measures for retrieval quality, an analysis of end-user questionnaires indicates considerable quality differences from search to search. Therefore, if the choice of the host can influence precision and recall, this certainly must be taken into account. Other factors like timeliness and type of implementation of a database, regional or local considerations, delivery time of offline prints, availability and reliability of the host as well as special features of the query language can be important for the choice of the host. The every-day practice of intermediary online searchers confirms that in addition to rational and financial considerations also psychological factors affect the individual decision for choosing a special system.

Introduction

Information business has changed from something like a very exclusive hobby for highly specialised people in the early seventies to a real information industry, where millions of Dollars, Francs, Lire, Deutschmarks, etc. are floating from country to country. Therefore it is of interest to look at the reasons why these millions of currencies are moving into the one or the other direction. This paper describes the viewpoint of two professional information services working as intermediaries for both internal and external users. Normally, such an information service has access to various different host systems. The reason for this is not obvious at the first moment but here are several explanations.

Frequently, a new information service starts with access to only one specific host. Even if this host is something like a large "supermarket" with numerous different databases, it may not be sufficient. For a specific query some database may be needed which is not available on this host. This is the first step to take the password of

* Revised version of a presentation held at the EUSIDIC Annual Conference 1985, Bath, England, 29–31 Oct 1985.

another host. On the other hand, the accessibility of the same database on different hosts will prove desirable. This redundancy ensures basically the permanent availability of intermediary services.

It must also be considered whether more than one query language can be managed satisfactorily by one single person. In the early years of computerised information it was assumed that retrieval experts are similar to computer programmers, who are supposed to be very good in only one single programming language. But in the meantime the situation seems to have changed. Today many online searchers are using regularly more than one or two hosts. Moreover, they also have fun using an "exotic" retrieval language on a completely different system. This is something like a joy-factor, which results in the positive effect of job enrichment.

Other reasons for using more than one host-computer will be discussed later as factors decisive for the choice itself. Therefore, it can be assumed that an information service normally has access to more than one host. In this case, after having selected a specific database fitting into a given query, the very next decision is the choice of the host where to perform the search (1)(2).

Cost factors

One of the most obvious decisive factors is the aspect of cost. But a serious calculation of average cost and an objective comparison between different hosts is difficult. First of all the daily fluctuations of currency exchange rates (especially important for the European point of view) affect these calculations distorting the results.

What other factors can influence the cost of a search? One of them may be the response time. If a command is performed on an expensive host with a response time of only a few seconds and the same command is performed on a cheep host with a response time of several minutes (this really happens sometimes), the final cost will probably be more or less the same.

Similarly, the comfort of the query language influences cost, since a short and effective formulation of a search strategy will save time and therefore save money (3). Another important factor is given by the splitting of databases. If a search has to be performed on a whole file which is split up into three or four different databases, it will generate higher cost than at a host where the whole file appears in one single database.

In addition, discounts are granted to various users under special contracts. A lot of other special rates charged by individual systems only again influence cost calculations. Such rates can be mailing charges, minimum billing rates per month or per quarter of a year, file entry fees, charges for special commands, charges for saved searches, etc.

Since the calculation of cost appears to be a really difficult problem, a different approach could be the interpretation of statistics. Figure 1 is a diagram indicating the mean cost per search for five different hosts. These statistics represent data of the information service at the Austrian Research Centre Seibersdorf in 1983. The cost indicated is independent of offline prints and the databases used, giving the overall mean values of the whole year for each specific host (4). It is quite surprising that the mean cost per search varies not more than 20% from host to

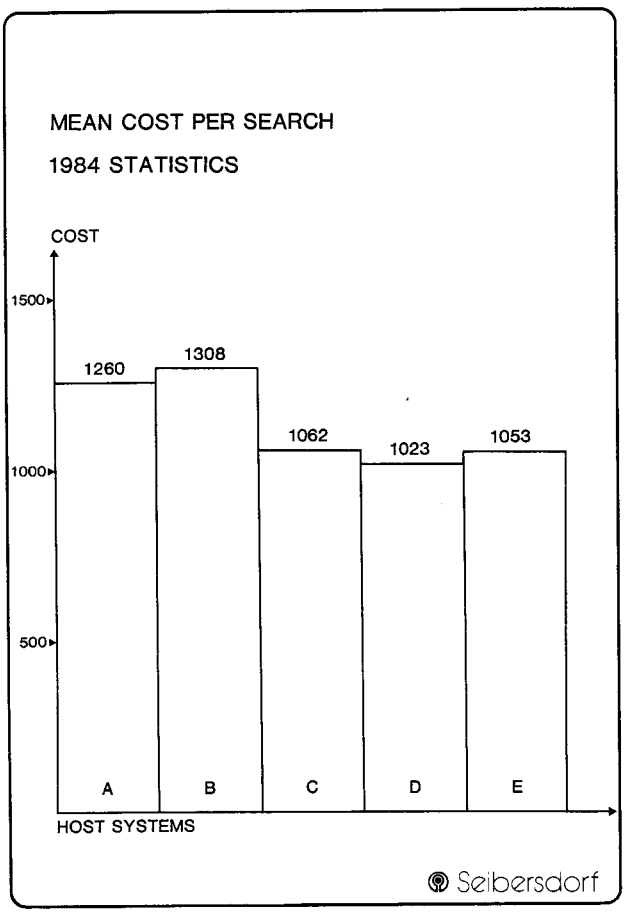

Figure 1

host. The hosts are two American and three European ones with a couple of different currencies. So we come to the conclusion that the factor "cost" is not really decisive for the choice of a host (5).

Search quality

If cost is less important for choosing a host, what other factors can influence this decision? The retrieval expert should never disregard the end-user for whom he has to perform the information services. The satisfaction of the end-user should be the ultimate goal for the way how to perform a search. Figure 2 shows the results of an evaluation of end-user questionnaires, in which respondents had to indicate the number of relevant/nonrelevant references (4). These results are split up by various databases. If precision and recall are accepted as measures for retrieval quality, this analysis of end-user questionnaires indicates considerable quality differences from search to search. The differences reach up to more than 100%. Therefore, the quality of a search represents one of the most important decisive factors for the choice of a host.

How can we improve search results? There are two relevant aspects: the materials and the tools to work with. The "material" is the database we want to search. In this context it is important how a particular database is implemented on various host systems (6).

One difference can be found in the searchable fields. It is almost incredible, but still true, that there exist hosts which have not inverted the abstracts for specific data-

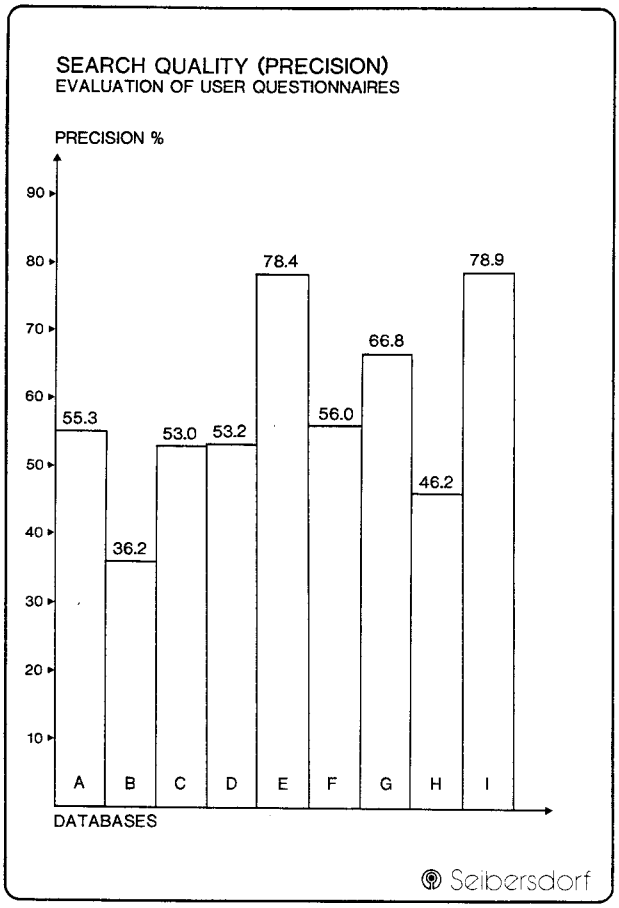

Figure 2

bases. One can really imagine that these circumstances could be a reason for avoiding the usage of such a host.

Another factor is the construction and the content of the "basic index". Some retrieval systems use to put all words of the whole record into the basic index. In this case one has to qualify subject-oriented search terms in order to distinguish them from formal-oriented categories. Other systems have only special indexes for the different categories. These differences of course affect the search (7).

In terms of time there are three aspects to be considered. First, timeliness can be a critical factor, if the purpose of the search is to identify the most recent information. A careful look on a particular database, established on several host computers, often shows considerable differences in the speed of processing the very last update. On the other hand, if the purpose of the search is to cover a period as long as possible, the host with the largest time range for the specific database will be chosen. Finally, the splitting of databases, even if increasing the search cost, can be advantageous for the search quality, if for instance a change of category codes causes a different search strategy from a specific date on.

Last but not least, the number and subject scope of databases available on a single host can be important for the choice of this host. Frequently, several different databases are to be searched for a specific problem. In this case one will probably prefer that host system where all the desired databases are available, so that one does not have to change the host when performing the whole query.

In addition to database implementation the query language (the above mentioned "tools" to work with) influences the search quality. In this context there are really significant differences. It is almost incredible, but even today there exist hosts which do not offer something like a proximity or context operator in the logic of their query language (8). It is evident that a retrieval expert will use such a host only if there is really no other system to perform this specific search. Truncation represents another point: On the one hand, there are many differences in the sophistication of the meaning of the truncation symbol. On the other hand, the maximum number of terms which is processed by a single truncation command is different from host to host. For example, it is really frustrating to enter seven times an asterisk in order to perform a truncation on a word stem identifying about 700 different terms.

Obviously a good retrieval language should contain the possibilities of field specification and qualification as well as limiting of terms. In addition, a subsequent field specification or a subsequent limiting (i.e. narrowing down a search strategy to specific subjects) represents a very high comfort of a query language. This reduces the cost of the query and on the other hand prevents the searcher from feeling frustration caused by wasting a lot of time on retyping the same words.

It is often underestimated that the print command (for online or offline printing of results) should be effective and easy to use. Also there should be convenient options for entering the end-user's name and the title of the query, as well as even a possibility for directing the offline prints to the address of the end-user. Although the latter facilities usually increase the total cost, the lack of such features may frustrate the searcher. And it is the searcher who makes the decision what host should be used.

Without aspiring to a comprehensive comparison of query languages, there are a couple of other features which are available on some systems, on others not, such as: individual adjustment of the software, online accounting information, its format and layout, document delivery service, SDI services, cross file searching, downloading support, etc. Therefore, if such a feature is necessary to perform a particular search, a system has to be chosen where this feature is available (9).

Comfort of the host system

The items considered so far usually determine the decisions for the individual search. Moreover, there are several factors influencing the searcher's attitude towards a host as a whole.

The first aspect is the user support granted by the host organisation, in form of newsletters, telephone or online. Although it is satisfying to have a telephone number for user support, any failure to answer calls will result in disappointment. If a help desk is offered, it should be available at least during the usual office hours.

Considering educational aspects, training seminars and user meetings: rational and commercial considerations are now to be enriched by human or social dimensions. It is desirable that the retrieval specialist feels part of a big user community. If a searcher has to choose between a faceless "black-box-computer" somewhere at the end of the line and a host system where he knows people – people using this host and those supporting the users of the

host – he will certainly prefer the system which gives the feeling of a human-human interaction.

Furthermore, the format and content of the host invoice could influence such a decision. In case a searcher has to make a very detailed accounting of the cost of the various queries, he really needs invoices indicating the cost broken down by date and time of usage and is disappointed if this information is not given. Another viewpoint of the "invoice-problem" refers to academic and governmental institutions using online services. Governmental institutions normally need a long period for settling of accounts. So it might happen that the invoices are balanced too late on part of the respective university or governmental institution. In this context unnecessary reminders produced by an unintelligent programme may cause a certain prejudice of the searcher against this host.

The format and layout of the offline prints can be important, especially for such information services which offer their results to external costumers paying for it. An attractive appearance of the results of an online search positively influences the searcher as well as the end-user.

Regional aspects

Concerning offline prints, the regional point also has to be taken into account. The postmark on the envelope of the offline prints will indicate whether the prints were processed on the day of the search or the day(s) after. Figure 3 shows overall statistics on the mean postal delivery times for offline prints, performed by three Austrian information services in 1984 (10). It is obvious that offline prints from a domestic host arrive much earlier than those from foreign countries, both from Europe or the United States. On the average, prints from American hosts need a few days more than prints coming from Europe. On the other hand, it is quite annoying if a particular print request unexpectedly does not follow this given pattern; the searcher will automatically blame the host for deficient processing.

Another regional aspect is telecommunication charges since increasing by distance. However, this factor is becoming less relevant because telecommunication charges on the whole have been reduced during past years.

In Europe the language barrier may be a decision factor for some specific queries – the language of the documents or of the search terms. But normally English is quite satisfactory as a carrier language for all information services.

Psychological factors

We must imagine that using an information system on a host computer is not a usual man-machine communication as we are used to think of the kind of interaction performed by computer programmers. Using an information system is similar to using a tool. The man sitting behind his terminal wants nothing else but solve a problem. Therefore he should feel comfortable in doing so. The query languages should be designed rather naturally than artificially. Simplicity is preferable, linearity is important.

The familiarity of users with a specific system causes a tendency for choosing just this system for their searches. One assumption could be that searchers tend to prefer that system which they learned first. In our opinion, however, the decision is influenced by some kind of inertia. Good experiences of a searcher with a host probably will cause repeated use of this system, bad experiences will evoke the opposite behaviour. Every retrieval specialist knows the annoying situation that the response time of a host is extremely bad. If this problem happens for several times at the same system, then the user will search for alternatives. Such and similar negative experiences can

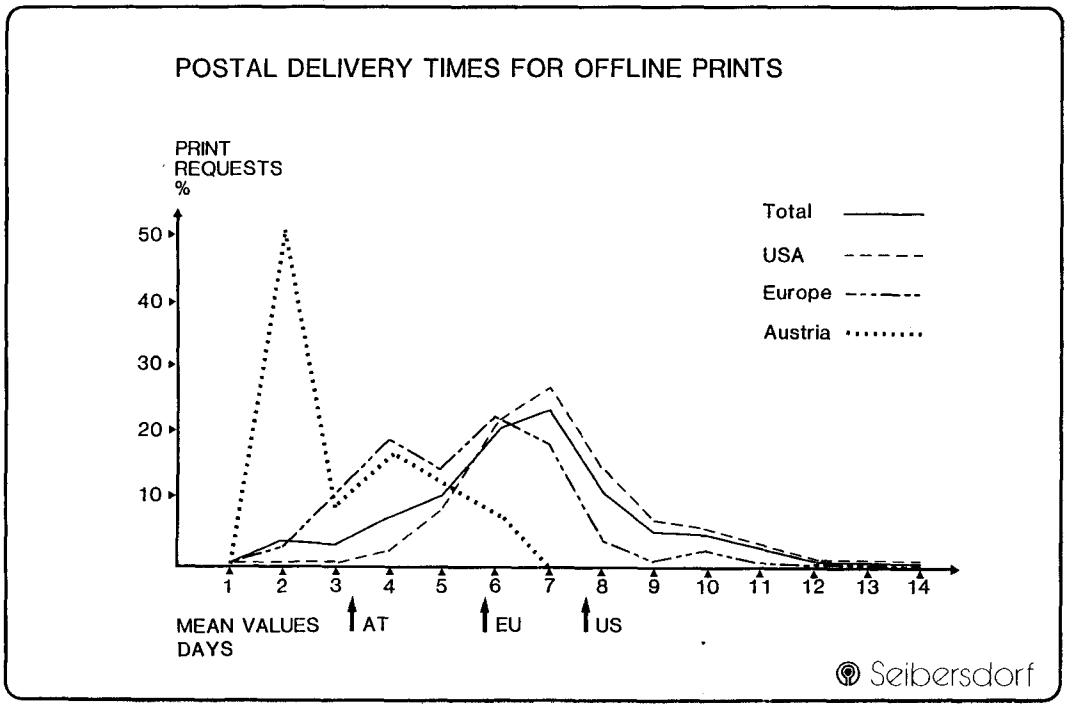

Figure 3

create averse predispositions which may last for long periods, even for years, regardless of possible improvements implemented meanwhile.

What is the first contact of a searcher with a new host? It will probably be the training seminar. So it is very important to perform exciting training seminars for those people who then will use this specific system. It is essential that the trainer who has conducted the seminar will be remembered as a sympathetic person not only enough experienced to perform this training seminar but also able to build up the first human-human relationship between the searcher and the host system.

Other personal contacts between users and host appear when calling the help desk. Large distances such as overseas inhibit the communication with the help desk because of telephone charges and time shift. Moreover, a (European) searcher will be pleased especially if his telephone partner is able to answer the problem in the searcher's own native language. Direct personal contacts by meeting host representatives for solving a problem will be rather rare – only if the host is situated in the same city as the searcher. On the other hand, user meetings, conferences, etc. are possibilities for such face-to-face communication, which from the psychological point of view have two major advantages: First, the everyday impression of the host system as a "black box" is replaced by real people individually reacting to the searcher's questions, problems, or even insults. Second, the group experience, with the host representative as a formal leader, can establish the impression of a "community" of users of the particular system improving familiarity and confidence.

This human factor is probably the only one which really can overcome all the prejudices and predispositions resulting from the frustrating and encouraging day-to-day experiences encountered with online systems.

Conclusion

Finally, the combination of all factors considered so far leads to the decision which host will be used for performing a specific search. The every-day practice of intermediary online searchers confirms that the individual decision of choosing a special system is influenced not only by rational and financial considerations but even more by psychological factors.

Acknowledgement

The authors are indebted to Mrs. Elfriede Pokorny for her editorial advice and the layout of the figures.

Adress of the authors:

Dr. Alexander *Nevyjel*, Österreichisches Forschungszentrum Seibersdorf Ges.m.b.H., Hauptbibliothek, A-2444 Seibersdorf, Österreich.

Dr. Otto *Oberhauser*, Universitätsbibliothek der Technischen Universität Wien, Karlsplatz 13, A-1040 Wien, Österreich.

References

(1) *Plosker*, G. R.; *Summit*, R. K.: Management of Vendor Services: How to Choose an Online Vendor. Libr. Managem. Bull., vol. 3, 1980, no. 4, pp. 2–5.

(2) *Halperin*, M.: Database Choice: A Practical Assessment. Ref. Libr., no. 5/6, 1983, pp. 187–194.

(3) *Hoover*, R. E.: A Comparison of Three Commercial Online Vendors. Online, vol. 3, 1979, no. 1, pp. 12–21.

(4) *Nevyjel*, A.: Enduser Feedback on Online Services: Concepts for Evaluation and Results from Austrian Users. Proc. DATABASE '83, Budapest, 6–8 June 1983. – Budapest: OMIKK 1984. Pt. II, pp. 628–643.

(5) *Burroughs*, B.; *Skaff*, J.: Maximizing Online Use on a Single System: Will it Really Save Big Money? Online, vol. 2, 1978, no. 2, pp. 55–64.

(6) *Oberhauser*, O.: Vergleichs- und Auswahlkriterien für "Chemical Abstracts" bei sieben Hosts. Fakten, Daten, Zitate, Jg. 4, 1984, H. 2, S. 13–17.

(7) *Conger*, L. D.: A Comparison of Basic Index Fields in Multiple System Searching. Online, vol. 4, 1980, no. 3, pp. 25–30.

(8) *Nevyjel*, A.; *Oberhauser*, O.: Gleiche Fragestellung – unterschiedliche Ergebnisse: Eine Fallstudie an fünf Online-Versionen der Chemical Abstracts. Online-Mitt. Komm. EDV-Anw. Verein. Österr. Bibliothekare, Nr. 16, 1983, S. 18–26. (auch OEFZS Ber. 4241; MI-57/83, 28 S.).

(9) *Conger*, L. D.: Multiple System Searching: A Searcher's Guide to Making Use of the Real Difference Between Systems. Online, vol. 4, 1980, no. 2, pp. 10–21.

(10) *Oberhauser*, O.; *Nevyjel*,, A.; *Hauffe*, H.: Postwegzeiten der Offline-Prints von computerisierten Literaturinformationsdiensten: Erfahrungswerte aus drei österreichischen Informationsvermittlungsstellen. Nachr. f. Dokum. 35 (1984) Nr. 6, S. 232–236.

Fachinformation zur Fahrzeugtechnik

Seit nunmehr über zehn Jahren wertet die Dokumentation Kraftfahrzeugwesen (DKF) die weltweit erscheinende Literatur aus dem Bereich der Fahrzeugtechnik und ihrer Randgebiete aus. Fachzeitschriften, Hochschulschriften, Dissertationen, Kongressberichte, SAE-Papers, Firmenveröffentlichungen – nahezu alles, was auf dem Gebiet der Fahrzeugtechnik veröffentlicht wird, wird in Ludwigsburg, dem Sitz der Dokumentation Kraftfahrwesen e.V., ausgewertet und in der Literaturdatenbank DKFL gespeichert, die mittlerweile 60 000 Literaturhinweise umfaßt.

Diese Datenbank ist seit Beginn 1986 nur für die Mitglieder der Dokumentation Kraftfahrwesen (Fahrzeughersteller, Zulieferer, Zubehörhersteller, Hochschulinstitute, Verbände und Behörden) online abfragbar. Über DATA-STAR wird die DKFL-Datenbank angeboten.

Neben der DKFL-Datenbank bietet die Dokumentation Kraftfahrwesen ihren Mitgliedern noch weitere Dienstleistungen im Informations- und Dokumentationsbereich: Eine Leistungs- und Konstruktions-Datenbank mit 300 technischen Daten und Messwerten aller in der Bundesrepublik handelsüblichen Personalkraftwagen, eine Testkartei mit dem Nachweis von Testberichten und Fahrzeugbeschreibungen (in Kürze ebenfalls online abfragbar), eine Sammlung der fahrzeugtechnischen Vorschriften und Richtlinien (ECE-, EG-Richtlinien, FMVSS-Standards etc.), den gedruckten Literaturinformationsdienst, Standardprofile und individuelle Literaturrecherchen.

DOKUMENTATION KRAFTFAHRWESEN e.V., Gröner-straße 5, D-7140 Ludwigsburg.