



Centre for
Agricultural Publishing
and Documentation

RESULTS OF A COMPARATIVE STUDY BETWEEN A MANUAL AND A COMPUTERIZED
CURRENT AWARENESS SERVICE

Paper presented at the European Regional Congress of Agricultural
Librarians and Documentalists (IAALD), Hamburg, 17-22 April 1978

J. van der Burg and H.C. Molster

July 1978

Box 4, 6700 AA Wageningen, The Netherlands



Centre for
Agricultural Publishing
and Documentation

RESULTS OF A COMPARATIVE STUDY BETWEEN A MANUAL AND A COMPUTERIZED
CURRENT AWARENESS SERVICE

Paper presented at the European Regional Congress of Agricultural
Librarians and Documentalists (IAALD), Hamburg, 17-22 April 1978

J. van der Burg and H.C. Molster

July 1978

RESULTS OF A COMPARATIVE STUDY BETWEEN A MANUAL AND A COMPUTERIZED CURRENT AWARENESS SERVICE¹

J. van der Burg and H.C. Molster
Pudoc, Wageningen, the Netherlands

Summary

The feasibility was studied of enlarging Pudoc's capacity to handle current awareness services by computer. Since 1954, the limit of the manual service was 40 profiles. The main questions were:

- 1) How much overlap is there between a computer service and a manual service?
- 2) How great is the lag of the computerized service behind the manual?

Pudoc's manual service is based on screening primary journals, and the computer service on secondary sources in machine-readable form. The main magnetic tapes used for the computerized service were CAIN (Agricola), Chemical Abstracts, FSTA, Medlars and Biosis Previews. The other two comprehensive agricultural databases CAB and Agrindex were unfortunately not available during the project. Only titles of publications were used that appeared in 1973 and which were found up till the end of 1974. The final results demonstrated the technical feasibility of a current awareness service for agricultural scientists based on existing computer-readable files and that the service can stand comparison with our manual current awareness service.

When all relevant titles, including "may be relevant", of the computer service and the manual service were totalled and duplicates removed, the manual service found 30% of all titles and the computer service 80%, with an amazingly small overlap of 10%.

Of all titles found manually, only 33% were also found by the computer; of those found by computer, only 12.5% were found manually. Thus the two services are generally complementary but when a choice is necessary, the computer service generally gives more than the manual service. However, these values are averages and individual profiles varied widely.

The computerized current awareness service is generally slower than the manual service, which is based on primary journals. Compared to our manual service, Agricola had a lag of two months, FSTA five months and Biological Abstracts eight months. However, Chemical Abstracts averaged two months faster than the manual system.

A comparison of the costs of the two systems will be discussed.

¹ Paper presented at the European Regional Congress of Agricultural Librarians and Documentalists (IAALD), Hamburg, 17-22 April 1978.

The figures presented in this paper are the result of a comparative study by Pudoc over several years. The aim of this project was to evaluate how and to what extent Pudoc could enlarge its current awareness capacity by utilizing computer retrieval techniques. The existing manual service then was limited by staffing to 40 profiles and the prospects for more staffing were bleak. The manual current awareness service has been running since 1954 and is manned by one graduate scientist, the "screener" who selects the relevant articles and by one administrative assistant. This service is based on daily screening of all primary journals that arrive at the Central Library of the Agricultural University, Wageningen, about 200 journals a day. In order to test the suitability of a computerized current awareness service a pilot project was set up, in which the retrieval performance of the then available major databases of interest to agricultural science were compared with the results of the manual service. An essential difference between the computerized current awareness service and the manual service is of course that the former screens secondary sources in machine-readable form whereas the latter screens primary documents. Please note that we did not compare tapes and secondary printed forms. That would have been the obvious comparison but was beyond the scope of Pudoc's situation. The main question was: Is the product of a computerized current awareness service as good as that of the conventional manual service?

To answer that question we investigated the following.

- a) To what extent does the computerized service produce the same titles as the manual one?
- b) What is the time-difference for delivering the same title between the two services?

The titles found for some current manual subscriptions were compared with those obtained by matching the same profiles against tapes with bibliographic data on agricultural research. For the statistics we took only titles published in 1973, that were retrieved up till the end of 1974. The results were compared for 25 topics, listed in Table 1.

Table 1. Profiles whose results were used for comparison between manual and computer service.

Title of profile	Tapes used ¹
1. Influence of airborne fluorine on plant growth	CAIN; CAC; BIO
2. Coccidiostatica: control of a protozoan pathogen important for poultry	CAIN; CAC; MEDL; EM
3. Nitrovin - use and effect of a growth stimulator on poultry and other animals	CAIN; CAC; EM
4. Use of aeroplanes in agriculture	CAIN; COMPENDEX
5. Processed potato products for direct consumption, like chips, crisps, croquettes, canned and dehydrated potatoes	CAIN; FSTA; CAC
6. Human diseases caused by potato consumption: do potatoes, infected with <i>Phytophthora infestans</i> (black spot) cause <i>spina bifida</i>	CAIN; FSTA; CAC; EM; MEDL
7. Oils and fats in potato processing (crisps, chips). Which chemical transformations occur at high temperatures; influence on end products, ration saturated/unsaturated fatty acids	CAIN; FSTA; CAC
8. Storage of potatoes - transport, handling, preservation, protection against sprouting	CAIN; FSTA; CAC
9. Edible mushrooms	CAIN; FSTA
10. Effect of radiation on RNA, DNA, chromosomes	CAIN; CAC; BIO
11. Tissue culture (plants)	CAIN; BIO
12. Culture of spinach	CAIN
13. Influence of natural gas on plant growth and origin, production and use of biological gas (methane)	CAIN; CAC
14. Intensive pig-rearing	CAIN
15. Economic integration in production (from feed and baby animals to consumer products)	CAIN
16. Animal destruction products (waste products of slaughter houses like bone- and bloodmeal)	CAIN; CAC; FSTA
17. Influence of air pollution on plant growth (excluding those in 1, 22, 23 and 24)	CAIN; CAC
18. Influence of natural and accidental radioactivity on living organisms	CAIN; CAC; MEDL; EM; INIS
19. Economic aspects of starch production (demand, supply, market aspects)	CAIN; FSTA
20. Starch production in plants	CAIN
21. Organization of agricultural research	CAIN
22. Influence of ozone on plant growth	CAIN; CAC
23. Influence of airborne sulphur on plant growth	CAIN; CAC
24. Influence of airborne lead on plant growth	CAIN; CAC
25. Sunflowers: culture and products	CAIN; FSTA

¹ CAC = Chemical Abstracts Condensates
 BIO = Biosis Previews
 MEDL = MEDLARS
 EM = Excerpta Medica
 INIS = International Nuclear Information System
 FSTA = Food Science & Technology Abstracts

In all 25 cases, CAIN, or should I say AGRICOLA, was used. This was, at the time of the study, the only comprehensive agricultural database, roughly corresponding to the printed version known as Bibliography of Agriculture. It takes a selection of journal articles and monographs that arrive monthly at the National Agricultural Library in Washington. Other major databases used during this study were Food Science and Technology Abstracts, Chemical Abstracts Condensates, Biosis Previews, Medlars and Excerpta Medica. The results are listed in Figure 1.

A			B		
COMPUTER 3274 (70)	M+C 470 (10)	MANUAL 939 (20)	COMPUTER 2223 (61)	M+C 470 (13)	MANUAL 939 (26)

Fig.1. Overlap between manual (M) and computer (C) service of relevant titles published in 1973. A: including "may be relevant"; B: excluding "may be relevant", Between brackets figures in percentages.

On general total of titles, the manual service covered 30% of all relevant titles found with both methods, including the category "may be relevant", and the computer service covered 80% of all relevant titles. If all "may be relevant" titles proved finally to be not relevant the figures would shift in favour of the manual service, but the computer service still had by far the best coverage. Presumably all manually retrieved titles were relevant, because not only the title but the whole document is available for assessment, whereas the computer screens only title fields, and descriptor fields, if available. Differences between the various subjects were rather large. Figure 2 demonstrates a frequency distribution of the coverage of both services.

In general, the coverage by the manual service is not very satisfactory, except one profile (80%). High coverages were found for five profiles (more than 90%) by the computer service.

Besides the number of relevant titles, the difference in

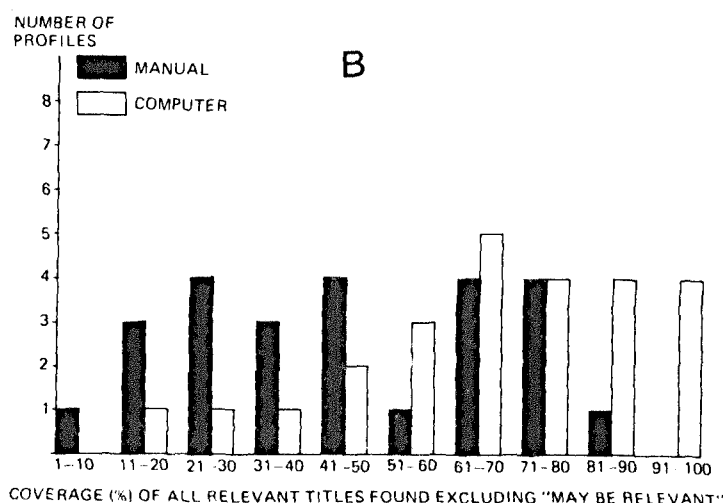
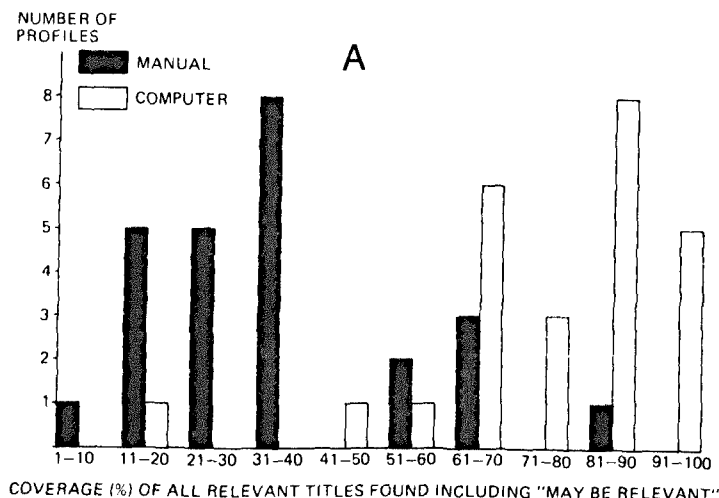


Fig.2. Frequency distribution of manual and computer service. A: including "may be relevant" titles; B: excluding "May be relevant" titles.

time-lag between the two services was also investigated as was the difference between the various magnetic tapes. The high speed with which the manual service alerts scientists to new literature is in Wageningen considered to be a great advantage: primary journals are screened the moment they are received at the University Library. The computer service is based on secondary sources which take some time to produce. Only those secondary services that make use of preprints of the primary literature can be faster than the primary publications themselves, for instance Current Contents. Another example of this feature is Chemical Abstracts. Various articles published in

Table 2. Average difference in time-lag (months).

	Manual	CAIN
Manual	0	+1.9
CAIN	-1.9	0
CHEM.ABS.	+2.1	+4.7
MEDLARS	-	-1.7*
FSTA	-4.8*	-1.7
BIOSIS	-8.1*	-5.7

1973 were already in Chemical Abstracts at the end of 1972. Table 2 shows the difference in time-lag between manual and computerized services and between the various magnetic tapes. All averages are based on 100 observations, except those marked with an asterisk, which are based on 40 observations.

The computerized current awareness service based on Chemical Abstracts is on average two months faster than the manual service, based on primary journals. A possible explanation is the availability of preprints and the fact that the Chemical Abstracts Condensates tapes arrive by air mail, whereas the primary American journals are sent to the library by surface mail.

In all other comparisons, the manual service was faster: 2 months faster than CAIN (AGRICOLA), 5 months faster than FSTA and 8 months faster than Biosis Previews. Because there were only a few observations for Medlars (about 20) only an estimate was possible: 4 months later than the manual service.

Figure 3 shows density distributions in percentages per month. The time it took for titles to appear in the various databases varied widely. It would be of much interest to know the delay in the CAB-tapes which provide, as do Medlars and FSTA, extensive abstracts. On the whole, we have little information on the importance individual users attach to the gain in time offered by the manual service. It would be of interest to know what delay is still acceptable.

If one compares two information systems one has to distinguish differences in retrievability and in "starting material". The "starting material" is the collection of the University Library on one hand and the contents of the tapes on the other.

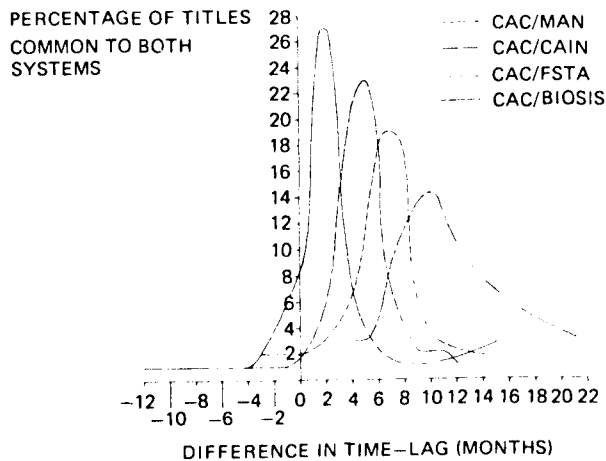


Fig.3. Density distribution of proportion of titles found by two systems as a function of the lag in time of retrieval from the second system with respect to the first. For instance, 27% of the titles common to CAC and manual were retrieved 2 months later by the manual service.

It was not feasible, given the time and staff available for the study, to find out why the titles found manually were not found with the computer. In many cases, titles not found with the computer service would be present in the databases but, because of the rigidity of the retrieval system, the vague titles, and because interest in special aspects of the article is often not expressed in the title, they could be missed, for instance if you were interested in methodology. For a number of profiles, our University Library has clearly a more suitable collection, namely for material of local interest. With the manual service, the chance of missing relevant titles is practically zero. The whole document could be screened if the manual searcher doubts. Of course, the starting materials differ in the various databases. Good knowledge of their contents is essential for selection of databases and formulation of profiles.

After this analysis of what is provided by one source and what by the other and what is the difference in delay between the services, it was necessary to examine the real costs of both systems (Fig.4). The costs per profile decreased as the number of profiles increased. Besides the costs of computer processing and the tape costs, we have also costs of staff. At the moment the computer service is run by one graduate scientist

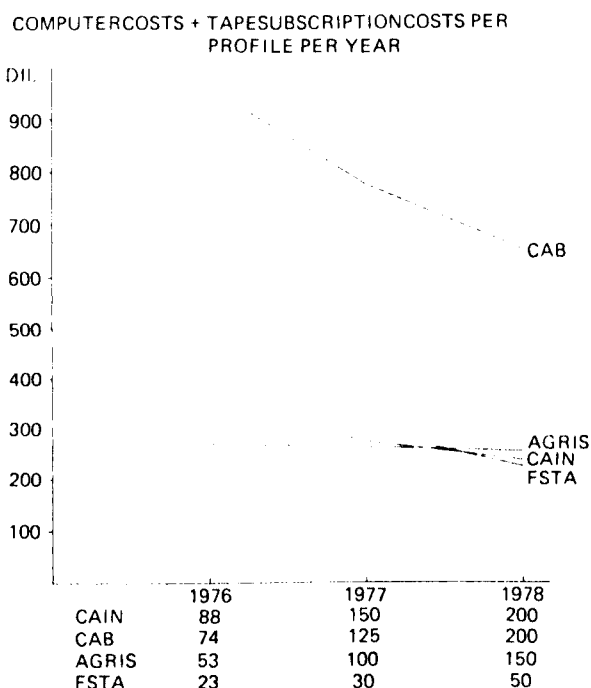


Fig.4. Computer and tape subscription costs per profile per year. The number of profiles is given per year (for 1978 the figures are estimated).

(full-time) and one administrative assistant (part-time), resulting in total staffing costs of Dfl. 101 000 per year. The staff could handle a maximum of 500 profiles, although the "man-power" needed depends more on new profiles than on total number of existing profiles, since the formulating and improvement of new profiles is the most time-consuming part of the operation. One can prepare at most 4 profiles in a day. In Fig.5, total costs are outlined. The computer costs consist of processing costs per database plus database independent costs (for example profile changes, paper costs), multiplied by a factor 2.11 for the average number of databases run per profile. We expect to run 200 profiles at the end of 1978. Total costs will then be Dfl. 281 000 or about Dfl. 1 400 per profile.

In the manual service one only has costs of staff (1 scientist and 1 administrative assistant) because the "starting material" is at Pudoc's disposal free of charge. Table 3 shows the costs of the manual current awareness service. With 200 computer

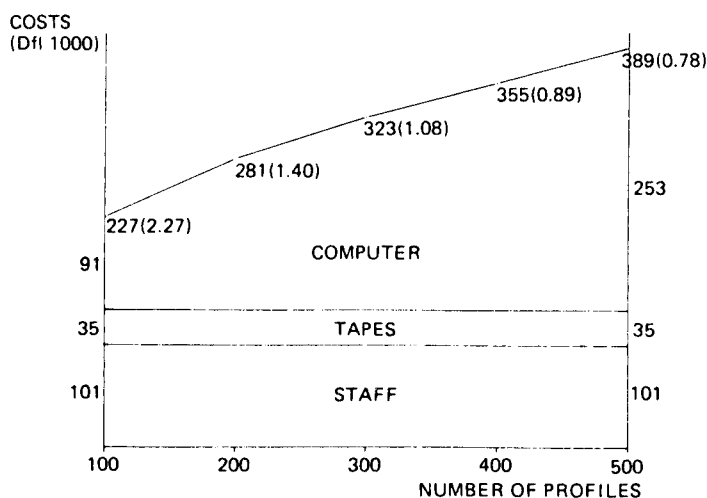


Fig.5. Total costs of the computerized current awareness service. Between brackets costs per profile.

Table 3. Costs of the manual current awareness service.

Number of profiles	Total costs manual service (Dfl 1000)	Costs per profile (Dfl 1000)
30	130	4.33
40	150	3.75
50	170	3.40

	Computer service	Manual service
Total subscription costs per profile (Dfl)	1400	3400
Costs per title (Dfl)	13	61

profiles at Dfl. 1 400 the client received on average 108 titles a year, which cost Dfl. 13 per title. Figures are derived from the results outlined in Figure 1, assuming that all "may be relevant" titles proved finally to be irrelevant. With the manual service the cost was at least Dfl. 3 400 and the client received on average 56 titles a year which cost Dfl. 61 each. The manual service is then about 5 times as expensive per title retrieved, when all "hidden" costs are included.

Pudoc's Current Awareness Service to-day: In the past, during the comparative study, AGRICOLA was the only comprehensive agricultural database, with FSTA as a database specialized in

Table 4. Number of profiles run against various databases.

Database	Number of profiles
CAIN	94
CAB	156
FSTA	47
AGRIS	96
CHEM.ABS.	56
BIOSIS	29
Total	478

food technology. Nowadays, the CAB database (Commonwealth Agricultural Bureaux) is processed as is the FAO database AGRIS. Processing of AGRIS and the equivalent of the veterinary journals of CAB (CABVET) started in 1975; the full CAB database has been processed since 1976. Current awareness services for Chemical Abstracts and Biosis are obtained from outside sources. As from March 1978, various databases are being searched with the following number of profiles (Table 4). Up till now we have recommended AGRIS only as a second database, because we do not think that it covers the world agricultural literature in a comprehensive way. Especially with the recent increase in the number of records from the United States, this situation is of course changing rapidly, although the contribution from some major countries is still lacking or insufficient. In 1975, 48 000 documents were put in AGRIS, in 1976 77 000 and during 1977 98 000. For 1978, the forecast is 140 000.

We charge a flat fee per database, independent of the number of terms per profile or output size. CAB is more expensive than AGRIS and CAIN, because of the higher costs of processing and printing of the abstracts in the CAB records. Since March 1978, we have been using the AGRIS database interactively on-line through the facilities of the International Atomic Energy Agency (IAEA) computer in Vienna within the framework of their INIS/AGRIS on-line experiment. We hope to use the retrospective AGRIS file intensively, in addition to the various files (CAB, AGRICOLA and FSTA) on the systems of Lockheed, System Development Corporation and the European Space Agency.

Conclusion

After a period of two years during which the results of both services have been compared, one can draw the following conclusions:

- (1) For most of the SDI-subjects manual current awareness service can be replaced by computer service; in most cases computer service will offer more relevant titles than the manual service.
- (2) The best average is obtained by a combination of the two sources, as both systems offer unique references.
- (3) Manual service is on average somewhat faster than the computerized service, but the difference is only a few months.
- (4) Availability of the original document is no problem for the manual service; photocopies are sent automatically on demand. Originals of references retrieved in the computerized service are sometimes inaccessible.
- (5) The manual service is on average five times as expensive per title retrieved as the computerized service.