A SURVEY OF THE CONTENT

AND CHARACTERISTICS

OF ELECTRONIC ABSTRACTS

Alan Wheatley

Department of Information and Library Studies, University of Wales, Aberystwyth, SY23 3AS

alw@aber.ac.uk

C. J. Armstrong

Information Automation Limited, Penbryn, Bronant, Aberystwyth, SY23 4TJ

lisqual@cix.compulink.co.uk

October, 1997

ABSTRACT

The study had three investigative areas: an examination of current database producers' guidelines for their abstract writers, a brief survey of abstracts in some traditional online databases, and a detailed survey of abstracts from three types of electronic database (print-sourced online databases, Internet subject trees or directories, and Internet gateways). The database producers, traditional online databases, and Internet databases were identified as representative of electronic information sources relevant to the higher education community in Britain, and were selected on the basis of both technical criteria and availability.

Abstracts were investigated to secure quantitative determinations of their properties in two broad areas. Their content was examined to ascertain the abstracts' coverage of source document concepts, to quantify their depiction of source document elements such as bibliographies, figures and tables, and to see if they acknowledged any geographical constraints of source documents affecting their value for users. To assess their physical and readability properties, abstracts were subjected to readability testing software that measured primary characteristics such as total length, sentence length and word length, and applied several standard readability tests.

Database producers' guidelines gave a clear view of the intentions behind professionally produced traditional (printed index based) database abstracts, and the brief survey provided benchmarks for comparing traditional and electronic database abstracts. These two studies allowed the validity of the detailed study of electronic abstracts to be tested at several points.

The detailed study found consistent but imprecisely defined differences between abstracts drawn from the three types of database. Those from print-sourced databases were the longest, but scored poorly in readability tests. By far the shortest and most readable abstracts were those from the Internet subject trees, and abstracts from the Internet gateways tended to hold intermediate positions. Internet gateways provided the most informative abstracts, closely followed by the print-sourced databases, whilst the shorter abstracts found in subject tree suffered the penalty of their brevity. Not surprisingly, abstracts from all sources made good use of their source item's title and first paragraph information, but the coverage of enriching features such as bibliographies, figures, and tables was disappointingly low.

Recommendations for further research include work to assess the relationships between measurable abstract properties and their value for users, closer examination of subject-related differences in abstract content, further work on the value of readability tests as a research tool, and investigation of metadata for improving the use of abstracts in the WWW environment. Recommendations for abstract producers include suggestions for the length of electronic abstracts, gaining a wider appreciation of abstract writing techniques used in all three types of database, considering the value of readability tests for quality control of abstracts, and maximising the information content of abstracts.

TABLE OF CONTENTS

1.0 INTRODUCTION

Once upon a time, every information worker knew that all abstracts had only one general form: they were commercially produced and printed on white paper, they were bound between hard covers, and their contents were concise and grammatically precise summarisations of larger items. ANSI's *American standard for writing abstracts*[1] defines them more generally and more usefully as "an abbreviated, accurate representation of the contents of a document." The diversity of modern electronic sources has bought us to a state where this is no longer the whole truth. Most importantly, the World Wide Web is a new networked environment for abstracts that allows numerous non-commercial originators to operate outside the constraints and quality norms of the commercial world, distributing documents encoded with the HTML formatting language to millions of potential readers. HTML coding is limited in its ability to faithfully identify and represent all the characteristics and structures of printed documents, but its capabilities are increasingly being extended by adding metadata, or data about its content, to the basic HTML document. Internet subject trees (Excite, Yahoo, etc.) and gateways (ADAM, OMNI, etc.) can exploit these metadata additions to invisibly simulate fields such as author and title that users from the "non-electronic" world already find familiar and useful.[2]

Networked information systems can deliver formal documents in the traditional mould, with or without enrichment by intelligent abstracting and indexing, and they can also provide a more diverse range of documents, such as personal Web pages, manufacturers' guide sheets, lists of links to other resources, and comments on formal documents. Few of these newer documents have their own abstracts, and few are likely to inspire abstract writing because of their non-commercial nature. The abstracting function for these newer documents must often be served by a diversity of newly developed "micro-texts" which may be inspired by, or automatically extracted from, a source document. Functionally, these micro-texts are equivalent to the abstracts of old. Some widely used examples of these new micro-texts are:

- Yahoo's one-line or two-line characterisations of extensive web sites.

- AltaVista's simplistic sampling of the first parts of web pages.

- Infoseek's more "intelligent" extraction of text from the body of pages.

- Broad summarisations of resource collections employed by information gateways such as ADAM and SOSIG.

The physical, intellectual, and operational characteristics of these new forms of document surrogates, and their relationships with the familiar abstracts of the printed index and online database world are largely uncharted. Mapping some of this unknown territory was the concern at the heart of this research project.

The project was conceived as an exploration of the significant "differences" seen in abstracts, especially those found in electronic information sources. Interest was centred on abstracts from online databases, commercial Web resources such as the general Internet subject trees, and the non-commercial cooperative subject gateways such as SOSIG and ADAM, but it was also necessary to establish some bench marks to serve as reference points.

---

[1] *Z39.14-1979: American national standard for writing abstracts*. ANSI: New York, 1979.
[2] Lorcan Dempsey. Meta Detectors. *Ariadne*, Issue 3. http://www.ariadne.ac.uk/issue3/metadata/

A major constraint on research was the need to complete all work in a two-month period. This was a most significant pressure, because it prohibited an iterative investigation (a natural response to exploring so many unknowns) whereby research normally proceeds through processes of progressive hypotheses, experiment, analysis of errors, and the Grail of experience. Other effects of the research's brief time span were that it was not possible to:

- Develop a framework for judging the likely value of abstract content, whether judgements were carried out by project staff or by real users.

- Train project staff to deal with anything more than basic value judgements of abstract content.

- Recruit and involve real users of abstracts.

- Investigate user expectations and needs.

Expressed in a more positive way, because of the controlling influence of the two-month period available for research, analysis, and writing up, the data gathering rôle of project staff was restricted to reporting essentially factual observations. Similarly, the manner of recording was essentially limited to Yes/No responses, checking multi-point scales, and recording readability test results.

Assessing the quality of source documents, in both paper and electronic form, has had ample attention from many workers,[3] but the quality issues for abstracts remain relatively unexplored. It is likely that the criteria for source documents will also be reflected in abstracts, but the literature is silent on the reality of this supposition.

Prominent among the usual criteria for source documents are matters such as authority, validity, currency, and uniqueness. All of these present difficult problems for a brief investigation of abstracts, because they:

- Require independent research into abstract content.

- Are notably subjective.

- Require significant staff training to ensure standardised collection of research data.

Thus the project "design criteria" required methods which all but eliminated the need for training project staff (eight students), allowed a useful number of abstracts to be processed, and minimised the effects of subjectivity. Effectively, this meant:

- Avoiding any techniques requiring extensive training or standardisation.

- Preferring test criteria that permitted simple *yes/no* appraisals rather than more sensitively scaled measures.

- Exploring subjective quality issues by means of unconstrained five-point scales and binary yes/no decisions.

- Controlling the effects of personal factors by distributing data collection across eight parallel project workers.

- Substituting standard readability measures (via word processors) for more realistic assessments of ease of use of the abstracts.

---

[3] For example a short list of criteria at http://www.vuw.ac.nz/dlis/courses/847/m2resevl.html or a much more detailed set of documents from the SOSIG team at http://www.sosig.ac.uk/desire/qlcont.html (this set arises from the team's need to inform participating contributors). These electronically oriented lists parallel those developed over many years for printed resources.

1.1 THE RESEARCH STREAMS

All these considerations led to several parallel streams of research investigating a number of the most accessible matters. A full investigation of a variety of abstracts would probably have two main areas as its "primary" stream:

1. Ease of use and benefits of abstracts to their users.

2. Costs and skills needed for creating abstracts.

But, a study of these two areas demands access to more than the information within the abstracts, and it could not be adequately investigated within the constraints outlined above. The first area was ruled out because its subjective nature requires access to real users, and the second ruled out because it requires access to commercially sensitive production data.

A more attainable line of enquiry was devised, within the constraint of requiring only data found in the abstracts or in abstract producer's own documentation, which allowed useful inferences on abstract quality and their likely value to users, and introduced the abstract producer's viewpoint. The areas in this stream are widely, if intuitively, acknowledged by abstract users and producers. They were identified by Tenopir and Jacsó, in an admirably clear article,[4] are discussed at some length by the ANSI abstracting standard,[5] can be found in the study of producer's abstracting guidelines included here (Section 2.0), and occur in more theoretical special purpose studies by Molina,[6] and by Fidel.[7] For an earlier generation, these points were established by important studies in the early 1960s: Borko and Chapman[8] very firmly identified the major aspects of abstract content, and the value of these contents was validated by Resnick's[9] slightly earlier study of relevance prediction from abstracts.

This secondary research stream was expressed in the following areas:

1.1.1 The extent of the abstract's representation of the document

Two sub-areas were recognised:

1. Presence in the abstract of themes selected from document texts.

2. Identification of enriching or special aspects of the original document and an indication of their presence in the abstract. These were:

> Bibliography
>
> Tables, formulae, and figures found in the document's text
>
> Geographical constraints on its coverage or likely relevance

---

[4] Carol Tenopir and Peter Jacsó. Quality of abstracts, *Online*, *17*(3), May 1993, p44-55.
[5] *Z39.14-1979: American national standard for writing abstracts*. ANSI: New York, 1979.
[6] Maria Pinto Molina. Documentary abstracting: toward a methodological model. *Journal of the American Society for Information Science*, *46*(3), April 1995, p225-34.
[7] Raya Fidel. Writing abstracts for free-text searching. *Journal of Documentation*, *42*(1), March 1986, p11-21.
[8] Harold Borko and S. Chapman. Criteria for acceptable abstracts: a survey of abstractor's instructions. *American Documentation*, *14*(2), 1963, p149-60.
[9] A. Resnick. Relative effectiveness of document titles and abstracts for determining the relevance of documents. Science, 134(3484), 1961, 1004-5.

1.1.2 The character of the abstract

Identification of a dominant character was the research intention, but abstracts of mixed character were also anticipated and were explicitly allowed for on the worksheets used for data recording. The distinguishing characters recognised were:

Informative

Indicative

Reviewing

Critical

1.1.3 The abstract's ultimate source

It was not assumed that abstracts would exist only in a professionally written form, workers were asked to indicate one or more of:

Author's original abstract

Indexer/abstractor/reviewer's specially written text

Abstract matches a recognisable portion of the document (title, first/last paragraph)

Electronic abstract matches a pre-existing print abstract

1.1.4 The abstract's readability qualities

For speed and convenience, this area was confined to the readability statistics provided by a word processor:

Basic measurements of the size of abstracts:

Number of characters present

Number of words present

Number of sentences present

Number of paragraphs present

"Complexity ratios" of the basic numbers:

Number of characters per word

Number of words per sentence

Number of sentences per paragraph

Established readability measures, widely used as "reading ease" indicators:

Proportion of passive sentences

Flesch Reading Ease

Flesch-Kincaid Grade Level

Coleman-Liau Grade Level

Bormuth Grade Level

1.1.5 Enriching qualities peculiar to the abstract

Indication of the abstract's source, such as indexer's initials or original author's name

Date of the abstract (especially significant for volatile Web resources)

A rating system indicating a judge's opinion of an abstract's likely "worth" to users, such as Yahoo's sunshades logo, Excite's LEP (Little Excite Person) ratings

Consistently "structured" organisation of content

1.1.6 Factors fitting or limiting an abstract to a specific audience

Use of jargon

Estimates of the suitability of the abstract's style to its likely audience

1.1.7 Explicit quality claims made by the abstract

Uniqueness

Originality

Comprehensiveness

Importance

(And any other claim for out-of-the-ordinary qualities)

1.1.8 Presentational issues affecting the abstract's ease of use

Most of these items were value judgement issues for which workers rated abstracts with five-point scales.

Clearly paragraphed organisation of content

Ease of reading (the student worker's estimate, not a formal readability score)

Layout of the abstract on screen

Typography

Use of icons

Use of backgrounds

Use of tables

Ease of navigation, where this was a relevant matter

1.2 METHODOLOGIES

1.2.1 Aims

The project's methodology was cautiously designed to explore the broad landscape of electronic abstracts not to survey its minor topographic details, and its aims were subdivided in pursuit of three quite different sub-areas of investigation.

1.2.2 The three lines of investigation

1.2.2.1 Database producer's guidelines for abstractors (Study 1)

An ancillary study was carried out to look at the guidelines given to their abstractors by a range of commercial database publishers. Working with eleven sets of database producer's abstracting guidelines, this study investigated:

- Aspects of abstract content the producers felt should be controlled.

- The degree of control producers intended for abstract production.

- Whether the findings of separate detailed examinations of abstracts were in accord with producer's general intentions.

1.2.2.2 Brief examination of online database abstracts (Study 2)

A short study of the basic readability characteristics of some 27 online databases, all derived from the traditional printed index environment, was undertaken to establish:

- The range of abstract variability in several different databases and subject areas.

- Norms for basic numbers such as abstract and word length and readability scores, so as to estimate the validity of results from the detailed study of abstracts.

This study included the LISA and ERIC databases, chosen because of their good reputation, their probable familiarity to the report's likely audience, and to provide a familiar reference point for the largest possible number of the report's readers.

1.2.2.3 Detailed study of Internet subject tree, Internet gateway, and online database abstracts (Study 3)

This was the main area of data gathering and analysis for the project. The principal areas of detailed enquiry were intended to discover the major properties of abstracts, as far as this could be done without involving real users of the abstracts:

- Abstract length
- Word and sentence properties
- Readability of the abstracts
- Sources of abstract information
- Abstract structure
- Ability of abstracts to reflect original item's content

- Treatment of common enriching features of the original items.

Most of the research time and effort was absorbed by this detailed examination of abstract content, gathering data representing commonly accepted basic quality criteria. Additionally, when both the abstract and the original article's text where available in machine-readable form, readability measures were extracted for possible article/abstract comparisons. (This proved impossible to accomplish within the available time.)

1.2.3 Sampling considerations

1.2.3.1 General guidelines for sampling electronic abstracts

The sparse literature on abstracts gives no guidance on sampling criteria for an investigation into abstract content and value, indicating a conservative sampling scheme should be preferred. Rather than collect detailed information from a few abstracts, the project sought to minimise the risks of unknown sampling errors by collecting the largest practicable samples consistent with a framework of abstract "types" determined by experience, the funder's interests, and the general IR literature. A first stage was to construct a framework of representative abstract "streams" based on easily observable gross concerns, such as (in no conscious order!):

- Data sources available to the UK higher education community through bulk purchase schemes such as BIDS.

- Abstracts specially written for electronic database use.

- Abstracts originally written for print and online database use.

- Abstracts which appeared to be primarily dependent on easily identified sections of documents, such as, title, and first and last paragraph, or some combination of these.

- Abstracts with visible gross differences in properties such as length and complexity.

- Commercial and non-commercial origins.

- A cross section of subject areas.

- Interest of the MODELS project and the eLib programme in subject gateways, such as ADAM, EEVL, OMNI, and SOSIG.

- The increasingly appreciated, and criticised,[10] abstracts and quality ratings provided by the WWW directories or subject trees, such as Excite, Magellan, and Yahoo!

- A reference set of abstracts whose worth is widely known in the higher education and IR communities: ERIC and LISA were thought appropriate.

- A reference set of abstracts from the established printed and online database world.

- All materials and databases investigated had to be immediately available; within the time allowed there was no possibility of obtaining materials from distant sources.

- Recent abstracts and items (within last 12 months) were preferred, to reflect current rather than historic practices of the databases.

---

[10] Alison McNab, Betsy Anagnostelis, and Alison Cook. Never mind the quality, check the bandwidth. *Ariadne*, Issue 9. http://www.ariadne.ac.uk/issue9/quality-ratings/

No attempt was made to investigate automatically generated document *extracts* used by the Web search engines, such as AltaVista, Excite, and Infoseek.

A further pragmatic consideration to add to this list was:

- Sufficient examples in each set of abstracts to permit basic non-parametric significance testing if this seemed appropriate.

- To minimise the possibility of bias caused by differences in personal skills or preferences, several project workers gathered each set of database abstracts.

1.2.3.2 The sample of database producer guidelines for their abstractors (Study 1)

This sample's purpose was to help determine some basic benchmarks for assessing the credibility of other findings.

The nature of the sample was largely determined by personal contacts and the fact that some producers were unable to supply guidelines within the time scale or did not have formal guidelines. Some eleven producers provided guidelines for a total of fourteen databases (see Table 1).

1.2.3.3 The samples for the main study of electronic abstracts (Studies 2 and 3)

The sampling considerations just outlined resulted in a framework with three strata:

Traditional print and online sources:
ERIC, LISA, ISI (via BIDS)

Commercial Internet subject trees:
Euroferret, Excite, Infoseek, Lycos Top 5%, Magellan, Yahoo, WebCrawler

Commercial and non-commercial Internet gateways:
ADAM, EEVL, OMNI, NetFirst, SOSIG

1.2.4 The size of the database samples

In each of the investigations of real databases (Studies 2 and 3), the target for samples was 30 satisfactory abstracts from each database examined. In most cases this was attained or exceeded, but not always, in which case the database concerned was excluded from analysis. Sometimes database samples were reduced during analysis, because only subsets of data were being studied or because unsatisfactory records were eliminated.

1.2.5 Conduct of the data gathering

In the absence of any satisfactory sampling framework, great reliance was placed on crudely randomised selection of abstracts, with the project workers being encouraged to use their discretion and library and information training (all had recently completed undergraduate or postgraduate information and library studies courses).

When the database permitted, each worker performed simple searches with appropriate search terms. Simple one-word searches were sufficient to generate useful output sets, and

workers were directed to include search terms of both general and specific subject value for each database searched. For example, search terms included: grammar, frog, rugby, Calvinism, anthropology, sucrose, vitamin, management, communism, and culture.

When search output was offered, every third item was examined for possible inclusion.

Some of the subject trees or directories demanded a different approach because their hierarchic presentation requires multiple conscious selection from on-screen menus, making them inherently difficult to randomise. Their presentation is so physically varied as to make universal instructions unrealistic, so searchers used the mouse blindfold, clicking as a pin into paper, or tried to base each selection decision on different criteria to the previous one. Both of these procedures must be seen as exploratory responses to an awkward situation.

Many items in the output were rejected as unsuitable (their number was not recorded). Candidate items were most commonly rejected because they had no valid or analysable abstract. Typically this meant that the item consisted solely of links to other resources, with little or no text to aid users in differentiating or evaluating the links. More flamboyant Web sites sometimes offered significant amounts of "abstract" text but presented it as scattered snippets with no unique beginning, middle, or finishing points, and with no preferred pathway for readers. Less predictable reasons for rejection included an unfortunate loading error in one of the databases, whereby a proportion of abstracts were truncated at 250 words. Other abstracts were rejected when they proved to be beyond satisfactory analysis with the word processor's readability tools.

Abstracts passing these tests were analysed using an appropriate work sheet.

1.2.6 Data recording

Workers were encouraged to record many aspects of information about each abstract, and a simplified yes/no response space was designed to encourage this. In practice, the variability of the material posed problems that might have proved entirely resistant to satisfactory treatment, even after extensive training. All the project workers were forced to make decisions which were influenced by the work they were doing, and all of them found it necessary to go beyond the strict scope of their instructions.

The reality of a yes/no decision space is that a "yes" response is usually much easier to determine than a "no" response, because the latter is often clouded by considerations of applicability. For instance, abstracts were screened for claims of originality and comprehensiveness, but the context and associations of some abstracts made it clear that such questions were inappropriate. In these cases it seems there was an understandable human reluctance for project workers to record a firmly negative "no" score when "not applicable" might have been a fairer and less judgmental response.

Readability data was recorded by highlighting and copying each abstract's text into a word processor, and running its grammar checker to generate readability data. Some pre-processing of the HTML documents was necessary. They contain many spurious line-feed/carriage-return characters, and these were removed before testing with the readability proofing tools. The readability results are available only as a screen graphic, so these were transcribed onto work sheets for later entry to a spreadsheet. It would be an exaggeration to describe all this as a process of mechanical simplicity and reliability, and some of the anomalous readability data may be attributable to processing errors at this stage, but the statistics can also be strangely variable and misleading in their own right (see Sections 3.7, 4.1.6. and the Appendix).

1.2.7 The readability measures

Readability of abstracts has always been seen as an important measure of their quality, and the convenience of dealing with machine-readable abstracts was a powerful argument for exploiting the ready availability of readability scores and data from modern word processors.

Mechanised measurements of readability offered the project an opportunity to penetrate inside the unbroken entity of the abstract, and were the only practicable way for a short-term project to make any internal analyses of abstract morphology and content.

Other workers have been attracted by similar circumstances. Tenopir and Jacsó[11] used the Grammatik grammar checking software to derive their readability scores. They noted some cautions about properly preparing text for readability tests, such as stripping out spurious carriage return/line feed characters, and the need for care in interpreting results. For instance, the software may regard an author's initial as a sentence (because it starts with a capital letter and ends with a point and a space). But overall they were clearly satisfied with their use of the tests.

In muted contrast, Hartley, Trueman and Meadows,[12] accepted the use of tests in broad principle but stressed the simplistic assumptions behind what are essentially ratios of common basic text properties, and they also noted many interpretation difficulties liable to separate true readability levels from the measured levels. Hartley and Sydes[13] echoed these reservations but thought readability tests can usefully contribute to comparison of bodies of similar text samples. In another article[14] they showed that, despite supposedly performing the same measurements and calculations, different readability programs gave different estimates of the readability of the same texts.

Nothing in the results of this project contradicts these earlier findings. The readability scores were enlightening, but they were also disturbingly varied, sometimes wildly erratic, and seemed often to show wide score differences for texts which are not easily perceived as being markedly different in character. Nevertheless, applying several different readability tests in parallel allows some mutual checking between test results, and the tests were vital in facilitating the comparison of several hundred abstracts.

---

[11] Carol Tenopir and Peter Jacsó. Quality of abstracts, *Online*, *17*(3), May 1993, p44-55.

[12] J. Hartley, M. Trueman, and A. J. Meadows. Readability and prestige in scientific journals. *Journal of Information Science*, *14*(2), 1988, p69-75.

[13] James Hartley and Matthew Sydes. Are structured abstracts easier to read than traditional ones? *Journal of Research in Reading*, *20*(2), 1997 (in press).

[14] James Hartley and Matthew Sydes. A thorn on the Flesch: observations on the unreliability of computer-based readability formulae. Paper presented to the 8th Writing and Computers Conference, Institute of Education, London University, September 1995.

1.3 THE PROJECT'S MAIN RESEARCH QUESTIONS

It is to be expected that the characteristics of abstracts selected from distinctively different sources will reflect the "temperaments" of these source databases, the abstracts extending in an array from extensive texts written for traditional printed indexes to snappy one-liners written for popular WWW guides. Within the research constraints outlined above, the project sought answers to a set of questions on these characteristics, and to questions prompted by current received wisdom on the quality and value of some of the Internet's information retrieval services.

Beginning with the most formalised abstracts:

1. Making up the first group of abstracts were those from ISI, ERIC and LISA, which were taken as a yardstick measuring widely accepted practices from the world of traditional printed databases. The former uses only author abstracts, and both of the latter were believed to make extensive use of their own specially written abstracts, though both may also make substantial use of an author's abstract when is seems suitable. Given the regard these databases usually receive: *are they sufficiently similar in their properties to reveal the foundations of this regard, and are these properties usefully summarised by the measures it was possible to make?*

2. The ISI citation index databases are known to use author's abstracts exclusively (this was also believed true of many other abstracting and indexing databases, but no documented foundations for the belief were known). The breadth and purity of ISI's sample of author abstracts provides research material for a core question: *are author written abstracts significantly different from the specially written abstracts favoured by ERIC and LISA, and what are the differences?*

3. The subject gateways examined appear to have a fairly homogeneous character. They are all subject specialised, rely heavily on volunteer contributors, have abstracts written by information specialists from allied areas of interest and disparate organisations. The four gateways, ADAM, EEVL, OMNI, and SOSIG are excellent representatives of this kind of Internet resource, and their availability, subject coverage, intellectual level, and UK origin makes them especially relevant to UK users in the higher education community. NetFirst, from OCLC, is essentially a commercial gateway for Internet resources, but otherwise has similarities of service with ADAM, EEVL, OMNI, and SOSIG. Word of mouth reports clearly indicate these gateways have been well received by their users, so the research questions here are: *which of their properties seem likely to correlate with the users high opinions, and how do these properties differ from typical abstracts converted from the print environment?*

4. Four mainstream Internet subject trees, Excite, Infoseek, Magellan, and WebCrawler have quite different backgrounds and development histories. Though they present four different faces to their users, working with them leaves the suspicion that their differences are only skin deep. A research question here was: *are they as truly similar as they first appear, and how do they differ from longer reviews?*

5. The Internet subject trees or directories are well established but one of them, the Lycos Top 5% database is clearly in a class of its own. This is currently Lycos' major database of reviewed Internet sites (the older Sites by Subject database has all but disappeared). Its reviews are conspicuously longer and often seem more informative than those in other Internet subject trees. When tested a prominent site rating system with clear

numerical scores distinguished it,[15] and the site has long been held in high regard. Because of this widely held good opinion, the research question is: *in what ways, other than gross length, do these reviews differ from others?*

6. Yahoo is a subject tree with a history long enough for it to be regarded with a degree of familiar affection by many Internet users. It was the first of the large scale Internet information retrieval sites, its development coinciding closely with the development of the World Wide Web. This history may be reflected in its short abstracts, which were first created as a non-commercial activity and later passed into the uncertain early commercial years of the WWW. Yahoo's reviewed site coverage is considerably larger than its rivals are, and it is the only subject tree to pursue hierarchical retrieval of its reviews with genuine flexibility (other subject trees allow searchers very little freedom in working with their subject trees). However its reviews are a byword for terse economy of words. They are conspicuously the shortest reviews from any reviewing site, yet the site is widely regarded as one of the best subject trees. Thus there is an important research question for the Yahoo data: *does the brevity of its abstracts markedly affect any of its properties?*

7. Euroferret, the recently introduced and humorously (?) named UK Internet database of reviewed sites is probably an order of magnitude smaller than its US rivals, and its abstracts are often shorter than these rivals. Its attraction is its specialised, if incomplete, UK coverage. With a smaller pool of (UK) source material, and a smaller core market than its US rivals, Euroferret must be under considerable pressure to produce its abstracts as economically as possible. This naturally prompts but one research question: *have these economic pressures revealed themselves in the test measures?*

8. There is a larger question implicit in these discussions of print sourced abstracts, Internet gateways, and subject trees. The same question also grows from experience of using these three different kinds of abstract-bearing electronic resources: *in what ways are abstracts from print sources, gateways, and subject trees differentiated from each other?*

9. Finally, it is important to explicitly state what this research did not undertake. There was no investigation of :

- Users subjective views of the readability of abstracts.

- Users preferences for particular styles or features of abstracts.

- The ability of abstracts to assist users in predicting the relevance of source documents.

---

[15] The Lycos Top 5% rating system was significantly changed very soon after these tests were carried out.

2.0 DATABASE PRODUCER'S GUIDELINES FOR ABSTRACTORS (STUDY 1)

2.1 Participating database producers and their databases

Of the twenty publishers contacted, eleven responded directly and OCLC passed our request to one of their database producers, the National Library of New Zealand. The Royal Society of Chemistry sent guidelines for three databases, so the project was able to examine a total of fourteen responses and eleven sets of guidelines. Bowker-Saur provided a copy of the abstracting guidelines used in producing LISA, making this the only database examined by the main survey for which guidelines could be studied. Two sets of guidelines are for Internet gateways stemming from the eLib Programme and examined in Study 3, the main survey of electronic abstracts: SOSIG and OMNI.

Other producers were contacted but were unable to assist, because they use only author abstracts, such as ISI and Elsevier. IEE (INSPEC) use author abstracts where possible but responded to our request with a page of their own guidelines. Some producers, such as RAPRA, have no written manual available.

Table 1. Producers and their databases

Producer	Database
SOSIG	SOSIG
OMNI	OMNI
Institution of Electrical Engineers (IEE)	INSPEC
Bowker-Saur	LISA
National Library of New Zealand	Index New Zealand
The Welding Institute (TWI)	WeldaSearch
American Psychological Association (APA)	PsycINFO
Royal Society of Chemistry	Chemical Engineering & Biotechnology Abstracts (CEABA)
Royal Society of Chemistry	Analytical Abstracts (AA)
Royal Society of Chemistry	Chem Business NewsBase (CBNB)
Commonwealth Agricultural Bureaux International	CAB

The length and style of the producer's guidelines varied considerably: The Royal Society of Chemistry produce three separate volumes, each of around 80 pages, covering indexing, classification, and abstracting, with the main section on abstracts being in the order of 10 pages. The Commonwealth Agricultural Bureaux have a 41-page manual dealing only with abstracts. Compared to this some sets of guidelines extend to only a couple of pages. The longer guides tend to include lists of standard abbreviations (especially chemicals, countries, etc.), plurals, companies, etc.

2.2 Abstract length

The guidelines revealed no accepted norm for abstract length, recommendations varied from "As brief as possible" to "No limit," and from 50 to 500 words. Producers prefer short abstracts for the obvious reason that length affects production and storage costs, and printing, binding, and distribution costs too when paper copies are produced. In no case did a producer suggest that abstract length more important than content and providing an adequate description of the document content.

Table 2. Producer's guidelines for abstract length

Database	Length
SOSIG	No limit
OMNI	No guidance
INSPEC	less than 200 words if possible
LISA	60-80 words
Index New Zealand	Four to five sentences. Be concise, do not use unnecessary words and avoid adjectives and adverbs
WeldaSearch	50-100 with maximum of 150 words
PsycINFO	960 characters (c.120 words) max. (shorter for case histories and program descriptions). 1,300 characters for reprints, errata, and APA journals
Chemical Engineering & Biotechnology Abstracts (CEABA)	"As brief as possible"
Analytical Abstracts (AA)	"As brief as possible"
Chem Business NewsBase (CBNB)	"Concise and to the point." Often two or three sentences, but sometimes more and seldom fewer
CAB	Consistent with length and importance of original: indicative a short sentence to 500 words and informative 100-250 words

Several other sources have advice on this matter. The American Standards Institute[16] suggests that for normal articles and monographs "an abstract of fewer than 250 words will be adequate." Tenopir and Jacsó's[17] opinion was that "average lengths of between 100 and 250

---

[16] *Z39.14-1979: American national standard for writing abstracts*. ANSI: New York, 1979.
[17] Carol Tenopir and Peter Jacsó. Quality of abstracts, *Online*, *17*(3), May 1993, p44-55.

words appear ideal." Reflecting another school of thought on this question, Borko and Bernier[18] suggested that abstract length be proportional to the material being abstracted, and proposed between ten and twenty per cent of the original.

Another line of evidence is seen in the separate brief examination of a sample of 27 mainstream abstracting databases undertaken for this project. These databases showed an average length of 114 words, but this figure varied considerably, with the smallest average length for a database being 19 words and the largest being 258 words. Standard deviations for the databases were as heterogeneous as their means, the typical standard deviation was about one quarter of the mean, or smaller, but it could be as much as two-thirds of the mean—so abstract length is a highly variable quantity, both within and between databases.

Twenty-six online databases comprise a healthy sample and, with the eleven database producers' guidelines, it helps to set norms for the project's considerations. Almost forty organisations from the printed and online establishment clearly feel that abstracts somewhat larger than a hundred words can acceptably meet most users' needs.

2.3 Abstract structure

Where guidance is given on abstract structure, abstractors were allowed to use only a single paragraph; in the case of SOSIG, the software only allows for a single paragraph and in no case is more than one paragraph expected. This is of particular interest in the light of discussion on structured abstracts where it is thought that the use of multiple paragraphs with headings make the assimilation of information easier for the reader.

Table 3. Producer's guidelines for abstract structure

Producer	Structure
SOSIG	Single paragraph only is possible
OMNI	—
IEE	Single paragraph
Bowker-Saur	Single paragraph
National Library of New Zealand	Single unified paragraph
TWI	—
APA	Examples are all single paragraphs
RSC (CEABA)	—
RSC (AA)	—
RSC (CBNB)	Examples are all single paragraphs
CABI	—

---

[18] Harold Borko and Charles L. Bernier. *Abstracting concepts and methods*. New York: Academic Press, 1975.

2.4 Repetition in the abstract of information from the source's title, etc.

The general rule seems to be that abstracts should not repeat information already available in the title or, in one case, in the descriptors and indexing. However, there is some difficulty with this as the abstract is expected to stand on its own and, as RSC point out, it is not a good idea to employ obscure phraseology simply to avoid such repetition.

Table 4. Repetition of information from the title, etc. in the abstract

Producer	Repetition of information
SOSIG	—
OMNI	—
IEE	Do not repeat Title information
Bowker-Saur	—
National Library of New Zealand	Use synonyms for descriptors. Do not repeat author's name. Should stand-alone and make sense without title or descriptors
TWI	Avoid repeating title words but do not be afraid to repeat words "as this helps when searching by computer"
APA	Do not repeat title information
RSC (CEABA)	—
RSC (AA)	"Obscure phraseology should not be used to avoid repetition of words found in title"
RSC (CBNB)	Should add something to the title, never rephrase it
CABI	No repetition especially of title. Self contained

2.5 Specific inclusions and limitations

All producers have something in their guidelines on what should be included in the abstract. For eight of the databases this is quite specific, and covers items such as processes, materials, purpose, major theme, and so on—normally the items specific to the database's subject area.

In several of the guidelines information on inclusion is divided into a number of sub-sections ranging over several pages. Headings include: Organisms and materials; Location; Dates and duration; Experimental techniques, design and methods; and Results. The Welding Institute is the only producer to mention the opportunity for adding value to the abstract by highlighting information that an article's author may not have made explicit.

Only SOSIG, the National Library of New Zealand, and American Psychological Association make a point of anticipating and noting specific search limitations that should be detailed in the abstract.

Table 5. Producer's guidelines for inclusions in the abstract

Producer	Inclusions	Limitations to be noted
SOSIG	Subject; source	Geographical; time
OMNI	Purpose; description; source; audience	—
IEE	Reflect content and approach; note new developments; do not include historical preamble; should not refer to body of paper	—
Bowker-Saur	Reflect content or retain flavour of original; if listing sections makes an adequate abstract, do so	—
National Library of New Zealand	Should provide key words not present elsewhere in record	Note time period and spatial boundaries for social science materials
TWI	Include processes, materials, tests, products. Types of welding process and steel should be mentioned even if author does not make them appear significant. Conference details where relevant	—
APA	Include purpose/hypothesis/major theme (including author's viewpoint), methodology, results, implications/ conclusions	Study location, subjects (with descriptors such as chronological age, mental age)
RSC (CEABA)	Guidelines indicate what to omit/avoid and what to include	—
RSC (AA)	Guidelines indicate what to omit/avoid and what to include	—
RSC (CBNB)	Summarise and describe contents; facts and figures. Guidance given on what to omit/avoid and what to include	—
CABI	Guidelines for inclusions such as: purpose, location, techniques, organisms, start and finish dates, etc.	—

2.6 Style and grammar

Considerations of abstract style (informative v. indicative, etc.) and of stylistic matters peculiar to the process of abstract writing (grammar, etc.) are often difficult to separate in the guidelines. Matters of style have few widely acknowledged rules, and the guidelines are not especially specific. In contrast, grammatical matters are more amenable to detailed treatment.

On the matter of abstract style, there is little to be found in the guidelines in the way of a general pattern. Despite its role as a gateway to quality material, SOSIG wished to avoid being judgmental in its abstracts, but OMNI emphasised evaluation in its guidelines.

Bowker-Saur indicated that abstractors should not be critical. Either an indicative or an informative style, or even a mix of both approaches, can be used in most cases. Only NLNZ specifically demand an indicative approach. The normal guidance is to "suit the article." The PsycINFO manual included information on, and examples of, the different types of abstract.

Table 6. Producer's guidance on abstract style and grammar

Producer	Abstract style	Tense and grammar
SOSIG	—	—
OMNI	—	—
IEE	Mix of indicative and informative is acceptable; informative preferred	Active or passive; past or present tense
Bowker-Saur	Indicative is sometimes acceptable	Natural English; punctuation and style notes given
National Library of New Zealand	Indicative. Some help given with style. Active voice	Some guidance, e.g. do not use the apostrophe with key words; also refers to New Zealand Style Book for punctuation, preferred spelling, and grammar
TWI	Style to suit content of article	Use clear rather than elegant language. Notes are provided on "words to use with care"
APA	Informative, indicative or mixed depending on subject material: examples given. Write in a way that sounds natural	Active verbs, third person. Guidance given on, for example, nouns and verbs agreeing. Use nouns and verbs, not adjectives and adverbs
RSC (CEABA)	Real sentences should be used. Short sentences are preferred	Detailed grammar and punctuation notes given
RSC (AA)	Not in direct speech and imperative mood should not be used	Detailed grammar and punctuation notes given. Tense should be consistent where possible
RSC (CBNB)	Avoid journalese and jargon – try for "reasonably good English." The first sentence is important and should introduce the subject material	Detailed grammar and punctuation notes given; not necessarily the same points as other RSC databases
CABI	As appropriate to original: informative, indicative or mixed. Objective and not critical. Easy to read in straightforward standard English; short sentences; avoid jargon etc; readability and lack of ambiguity are paramount	Past tense for informative; present for indicative; active voice preferred

Grammatical guidelines were most in evidence from The Royal Society of Chemistry. They make special note of incorrect and correct usage of words, such as while/whereas/but, fewer/less, which/that, and may/can. Their guidelines include many such specific notes, such as not using apostrophe-t (as in can't and isn't), and not comparing the incomparable (as in "The results were compared with those of last year" instead of the correct "The results were compared with last year").

2.7 Spelling, and the use of abbreviations and numbers

Interestingly, neither of the Internet producers offer guidance in any of these areas, while the conventional abstract/database producers do so in nearly every case. The spelling guidance naturally reflects the origin of the database and, in the case of the RSC, presumably either its likely use in North America or common practice among chemists internationally.

Table 7. Producer's guidance on spelling, and the use of abbreviations and numbers

Producer	Spelling	Abbreviations guidance	Numbers guidance
SOSIG	No	No	No
OMNI	No	No	No
IEE	British or American but consistently and correctly	Yes	No
Bowker-Saur		Yes	Yes
National Library of New Zealand	British	Yes	No
TWI		Yes	Yes, but only for steel compositions
APA	American; use Webster's New Collegiate Dictionary	Yes	Yes
RSC (CEABA)	British (except for "sulfur" and –ize endings which are preferred)	Yes	Yes
RSC (AA)	British (except for "sulfur" and –ize endings which are preferred)	Yes (very detailed)	Yes
RSC (CBNB)	British (except for "sulfur" and –ize endings which are preferred)	Yes	Yes
CABI	British with some exceptions noted	Yes	Yes

In many cases the guidance given for abbreviations and use of numbering is very detailed. Abbreviations are particularly important in technical descriptions and this is reflected in the tables given by RSC, CABI and TWI (WeldaSearch). Also under the general heading of

abbreviations can be found guidance on words to be used with care, plurals, scientific names, and conventional usage.

2.8 General guidance to abstractors

Most producer guidelines also offer general help that is too inclusive to be easily categorised. The examples of such help mentioned in the table below are intended as indicative pointers rather than comprehensive summaries.

Guidance over the use of author abstracts is an important topic covered by such general notes, and the instructions vary considerably. IEE, Bowker-Saur, NLNZ, APA, and by implication, TWI, accept the use of some version of an author abstract, as do Elsevier and ISI (excluded from analysis because they use only author abstracts and so have no guidelines of their own). RSC explicitly indicate that an author abstract will not be adequate. CABI makes no comment on the use of author abstracts. SOSIG and OMNI also make no comment but, for gateways, it is not normally possible to use an author abstract.

The strictures and permissions relating to the use of entire author abstracts are sometimes extended to extracts from them. Bowker-Saur (LISA) encourages their abstractors to use extracts from the original, RSC indicate that this should be avoided. RSC and IEE both advise against extracting from the text, while Bowker-Saur indicates that this is a useful approach. Bowker-Saur does not regard this as plagiarism whereas RSC clearly does.

The RSC Guides also include an introduction to the general principles of abstracting, as well as advise on how to make good use of the source document. In another guide, Bowker-Saur includes a five-step methodology for abstracting (from "read the original" to "read through the abstract, clarify any obscure passages by comparison with the original paper and ensure that any figures given are correct"). Several producers wanted their abstractors to follow the overall structure and order of the original article.

Table 8. General guidance to abstractors

Producer	General guidance notes
SOSIG	Avoid being subjective or making judgements.
OMNI	Guidelines on where to find information, e.g. FAQs; emphasis on evaluation criteria
IEE	Author abstract used if available. An abstract should not simply be a direct selection of sentences from the author's introduction or conclusion
Bowker-Saur	Do not be critical. Use original abstracts and indicate this in field 500$b. Extracts from the original can be useful and, in general, it is not necessary to make changes
National Library of New Zealand	Use author's abstract unless it is very long. Indicate the major points and make general statements about document. Retain basic information, order and tone.
TWI	Authors abstracts may be re-written or replaced completely if inadequate.
APA	Journal abstracts may be used. Helpful tables (such as condensations: use "several" rather than "a number of") are given.
RSC (CEABA)	Section on Plagiarism versus Convergent Phraseology (see AA, below). "The reader should not have to ask 'Why has this item been selected?'"
RSC (AA)	Clear and concise and bring out what is new in the article. The abstract required for RSC will seldom be identical with that found in the original article. Avoid using phrases and whole sentences of verbatim text.
RSC (CBNB)	The abstract required for RSC will seldom be identical with that found in the original article. Avoid using phrases and whole sentences of verbatim text.
CABI	Contents must follow a logical sequence and be in the same order as the original.

2.9 Summary of the first study's findings

Conclusions must be drawn very carefully from such a small survey. It provides useful detailed accounts of current practices, but the variety of the accounts suggests that there are currently no common standards, and very little agreement on the guidelines given to abstractors.

The guides made available for the survey varied from extremely short to lengthy documents with tables of abbreviations, etc. However, the producer's guidelines are not the whole story. It should be remembered that, in addition to a written manual, training is normally given. The National Federation of Abstracting and Indexing Services (NFAIS) survey of 1994-5 indicated that formal training in indexing and/or abstracting often continues for three to six

months, depending on the technical nature of the material[19]. Work is also closely reviewed after the training period until the supervisor is happy with the standard reached. Most of the larger companies (such as the Information Access Company (IAC) or the Chemical Abstracts Service) visited during the NFAIS survey also have quality checking/assurance procedures. In the case of IAC this included both positive feedback for a well-written abstract and the marking of records containing errors.

---

[19] National Federation of Abstracting and Indexing Services. *Report on the Benchmarking Project*. Philadelphia: NFAIS, 1995.

3.0 BRIEF SURVEY OF ONLINE DATABASE ABSTRACTS (STUDY 2)

3.1 Introduction

It is important to appreciate the rôle of this brief study within the larger examination of electronic abstracts. Expressed concisely, this investigation of abstracts from online databases sought to summarise contemporary practices from current printed indexes and the commercial online databases generated from them. It records the readily quantifiable properties of abstracts produced in the last year or two, and provides some basic yardsticks to use in assessing the electronic abstracts.

Because this part of the study is based on a sample of about thirty records from each of twenty-six online databases (almost 750 records in all), generating summary figures can lead to a perilous walk onto the thin ice of statistics. It is improper to calculate ratios such as averages and medians for each database, and then to take averages and medians again to produce figures for the whole sample. Thus the only figures presented are basic averages and medians for the entire sample; all detailed figures for individual databases will be ignored.

3.2 Abstract length

The arithmetic mean length of all the records examined was 118 words. It was common for individual records to depart considerably from the mean; the standard deviations of abstract lengths usually fell between 25 per cent and 50 per cent of the mean length.

Given the generous span of the standard deviations, implying a considerable number of outliers among the individual records, the fact that the median abstract length was only 94 words is not so much a surprise as a reminder of the median's superiority for summarising data with many outlying values.

3.3 Abstract structure

Little of real value can be said under this heading. Visual inspection suggested that abstracts were normally just one paragraph long—a perfectly reasonable format for texts of about a hundred words. Word processor readability tests in reported the abstracts as having a mean length of nearly 1.4 paragraphs, but closer inspection showed the figure is very prone to distortion by the test's unsatisfactory ability to distinguish paragraphs, and it was dismissed as an untrustworthy artefact of the software.

Tough formally structured abstracts were not a significant feature of this study's sample, a number of the databases showed what might be called informal structuring. This occurred sufficiently often to suggest that abstractors were following explicit in-house practices. However, none of the databases consistently demonstrated a high proportion of structured abstracts. It seems that formal structuring is essentially a response to particular circumstances, types of article abstracted, author practices, etc.

3.4 Complexity of the abstract's words

In the records examined, the mean length of words is 5.6 characters, and the median length is 5.4. Again, the difference between the two figures results from the effects of data outliers, and should not be seen as overly significant. More interesting is the fact that the abstracts' word length shows less variation than their overall length: the standard deviations of the word lengths in individual databases usually ranged only between 7 per cent and 15 per cent of the mean length (25 to 50 per cent for abstract length).

Genuinely authoritative benchmarks are hard to establish, but among printers and publishers it is commonly reckoned that the average word length in a run-of-the-mill English language text is five characters, and the words in scientific, technical and medical texts are six characters. It is unsettlingly counter intuitive that a sample of abstracts representing (largely) scientific, technical and medical texts should have words noticeably shorter than the words reputably comprising these texts. Whilst it is tempting to speculate that this agreeable shrinkage of words represents the effects of abstracting skills, the high incidence of author's abstracts in the samples suggest otherwise. It seems the language of abstracts is not a fair sample of the language of source documents.

3.5 Sentence lengths

An appealingly simple statistic of the abstracts examined for benchmarking is the average number of sentences per abstract. Some quick tests on a random selection of material did not show the readability tests were able to determine the number of sentences per abstract with absolute reliability (sometimes the figure was over stated). Thus the readability tests' average, of 25 words per sentence, should be treated as a reasonable estimate rather than accurate report.

3.6 Passive sentences

A preponderance of passive sentences is widely regarded as a hindrance to readability, though the evidence advanced for the belief is not as clear as the strength of condemnations might imply. Predictably, the rate of occurrence of passivity is one of the readability measures generated by readability tests, and this incidence can be calculated as about 37 per cent (mean value) or 33 per cent (median value). Behind these overall figures are huge variations among the abstracts—from zero to one hundred per cent in many of the databases. In fact, the standard deviation was normally greater than half the mean, sometimes as large as the mean, and sometimes as much as twice the mean! Such variations make for bad descriptive statistics: the extreme values disturb calculation of the mean, and the clustering at integer values disturbs the median. Sustained inconsistency at these high levels necessarily reduces the diagnostic value of these measures, and they must be treated with caution.

3.7 Formal readability scores

The readability scores below were obtained from the aggregated scores in the hope of producing robust benchmarks to certify the detailed test scores in Study 3. Since all these readability measures are essentially ratios of word lengths and sentence lengths it was anticipated that there would be some statistical problems to overcome before arriving at useful benchmarks. This was a sad and inaccurate assessment of the situation.

Questions about the accuracy of the readability analyses will be discussed in more detail later, but it is here that the project encountered them for the first time. Quite simply, the analyses generated such varied scores that averaging them across a large body of data did little to massage them into convincing benchmarks. It may even be that the large body of data contributed to the inconsistencies by introducing a wider range of variations in the test material than the readability tests were designed to analyse.

Table 9. Readability scores for the aggregated online database abstracts

Readability tests	Mean	Median
Flesch Reading Ease	33.5	33.7
Flesch-Kincaid Grade Level	13.5	13.5
Coleman-Liau Grade Level	20.7	15.6
Bormuth Grade Level	11.0	11.1

Thankfully, the mean and median scores were in good agreement (apart from the inflated mean Coleman-Liau Grade Level score). It seems probable that whilst either way of summarising readability test scores is reasonably satisfactory within a given body of test material, but the mean scores are safer. Though the scores for each test are numerically different, they all concur in interpreting the abstracts as being at undergraduate level of readability. Flesch Reading Ease scores have anomalously high figures because they are calculated against a different scale to the other three tests. The three "grade level" tests all purport to report their scores in terms of American school grade levels, and these can be conveniently interpreted by adding five [years] to a grade score to convert it into an equivalent reading age. Bormuth scores were unexpectedly and consistently low compared to the other tests, and this was also observed in the detailed tests (Study 3, below).

3.8 Conclusion

At the benchmarking level, Study 2 proved very useful. Using a sample of conventional online database abstracts, it established four norms for important internal properties of abstracts: overall length, word length, incidence of passive sentences, and readability test scores. These norms were later found valuable in demonstrating the extent to which the same measures revealed useful properties of the electronic abstracts examined in the project's main tests.

4.0 DETAILED SURVEY OF ELECTRONIC ABSTRACTS (STUDY 3)

4.1 PHYSICAL AND READABILITY PROPERTIES OF ABSTRACTS

Given the reservations already expressed over the readability tests, the report will assess readability from several different perspectives, using simple measures of abstract morphology as well as the readability tests proper.

4.1.1 Abstract length

The hypotheses examined here are an expression of four broad premises:

- The cost of creating abstracts is directly related to their length, and is normally minimised, but abstracts created by authors or for non-commercial databases are not so directly influenced by this pressure.

- The additional costs associated with printing, binding and distributing print sources creates pressures to minimise their length.

- Commercial imperatives require that creation costs be minimised, as long as this is consistent with the value likely to be perceived by users.

- Within the limited range of abstract lengths seen in these tests, longer abstracts are probably more informative and more desirable than shorter ones.

These premises suggest three related hypotheses. The longest professionally created abstracts are likely to be found in the most widely used databases, the shortest abstracts are likely to be created for new or small and specialised databases, and users will prefer the most informative [longer] abstracts (this hypothesis not testable within this test's confines).

The enormous range of abstract lengths seen in the table below is startlingly obvious, and points very clearly to producers holding very different views of user and subject needs. Perhaps this, or an accident of sampling, is the explanation of the unexpected fact that the median length of the online database abstracts—94 words—is larger than two on the online databases included here, and larger than all but one of the remaining databases.

Within the table's broad picture, five points emerge, providing partial support for the hypotheses.

Table 10. Median number of words per abstract

Database	Abstract length	
ISI	136	*(best)*
Lycos Top 5%	118	
LISA	92	
EEVL	69	
ADAM	66	
NetFirst	60	
OMNI	58	
SOSIG	58	
Magellan	53	
ERIC	51	
Excite	46	
WebCrawler	45	
Euroferret	32	
Infoseek	31	
Yahoo!	17	*(worst)*

In this and all the other tables summarising test results, the following display scheme has been used:

Yellow	*Internet subject trees*
Red	*Internet gateways*
Blue	*Print-sourced databases*

1. The abstracts provided by the Internet subject trees and gateways were markedly shorter than abstracts *typically* provided by traditional online databases such as ERIC, ISI, and LISA (though ERIC's abstracts are much shorter than ISI's and LISA's). Averaging the figures in table 10 gives abstract lengths of about 54 words for all Internet sources (49 words for subject trees and 62 words for gateways), and 93 words for the print sourced databases (compare with 114 words for the online databases examined in Study 2.

2. A line drawn across the table at about 56 words (the "average" of these medians) bisects the data. Below it are six of the seven commercial Internet subject trees, above it are all four non-commercial Internet gateways, and the remaining commercial sources. It is abundantly clear that commercial pressures are a very strong determining influence on the length of abstracts, but that other factors are also at work.

3. The longest abstracts are the (no-cost) author abstracts found with about half of the items in the ISI databases.[20] Most unexpectedly, in second place are Lycos Top 5% abstracts, which are of surprising length (more than twice the size of reviews in any other subject tree), perhaps accounted for by Lycos' aggressive pursuit of users, its commercialism, and its avoidance of printing and binding costs. This length might be a factor in Lycos' enduring popularity. LISA's abstracts perhaps reflect their generally perceived quality.

---

[20] Bonnie Snow. [Caduceus column] SCISEARCH changes: abstracts and added indexing. *Online*, *15*(5), September 1991, p102-6.

Interestingly, this well marked group three databases with the longest abstracts includes two originating as printed sources.

4. The large group of abstracts with lengths between 45 and 69 words is a diverse one, and surprisingly it includes ERIC (the only print-based product) somewhat below its middle point, and well separated from the other print sourced databases. NetFirst and Magellan, the other commercial databases in this group have longer abstracts than ERIC does. The consistently high placing in the table of all the Internet gateways correlates with the project workers informal expressions of their quality, and Web Crawler's place at the bottom of the group also coincides with their informal quality comments.

5. A disproportionately large gap separates Euroferret and Infoseek from the middle group of abstracts. Again this is in line with informal quality comments: several project workers described them as "thin."

6. Given its markedly economical abstracts, Yahoo's position at the foot of the table is inevitable, but this placing is not a reflection of the site's popularity or high reputation. Perhaps the discrepancy is accounted for by factors such as the excellent organisation of its reviews and the site's long-held good name over-riding the limited scope of its abstracts.

7. An observation spontaneously made by several project workers was that the Internet subject trees' short abstracts were initially appealing, but the appeal lessened with growing experience of longer and more informative abstracts in other databases. Gathering information on subjective preferences for different forms of abstract was outside the scope of the research design, but these observation appear significant.

A hypothesis of print production costs negatively controlling abstract length seems entirely unsupported. The strength of commercial pressures is firmly confirmed. A general association of abstract length with user perceptions of quality and value appears to have strong anecdotal support.

4.1.2 Complexity of the abstract's words

The hypothesis here is an extension of the generally direct relationship between short words and a text's ease of reading. Word length (in characters, excluding the space) is a widely used basic measure of the ease of reading of words, shorter words being taken as indicative of more readable text. There must always be reservations about the measure, because:

- A word's familiarity to users can override difficulties caused by its length, and *vice versa*.

- Some source articles and subjects require more complex words than others do.

- The readability tests do not distinguish adequately between words and other short sequences of characters, such as acronyms and initials.

The hypothesis is that word length will be shortest in the more consumer oriented WWW subject trees (Excite, Euroferret, Infoseek, Lycos Top 5%, Magellan, WebCrawler, Yahoo), longest in the abstracts written for professional audiences (ADAM, EEVL, ERIC, ISI, LISA, OMNI, SOSIG). As the only pay-per-view subject tree, written with professional needs in mind, NetFirst seems likely to fall between these camps.

Table 11. Median word length of abstracts (in characters)

Database	Word length	
Lycos Top 5%	4.80	*(best?)*
Excite	5.00	
Infoseek	5.10	
Magellan	5.20	
WebCrawler	5.30	
OMNI	5.40	
Euroferret	5.45	
Yahoo!	5.50	
LISA	5.60	
ISI	5.60	
ADAM	5.60	
SOSIG	5.70	
NetFirst	5.70	
EEVL	5.80	
ERIC	6.05	*(worst?)*

Word lengths in the table effectively bracket the median value of 5.4 characters observed in the online databases (Study 2). The table's data strongly support the hypothesis, but there are some minor surprises in the ranking. With the notable exception of OMNI, all the professionally oriented databases tend to use longer words, and to rank below the popular consumer oriented databases. ERIC and Yahoo's low rankings (within their groups) in the previous table are counterbalanced here by the demonstration that the shortness of their abstracts may be partially compensated by employing words longer than the words in other subject trees and thus of a greater information carrying capacity.

4.1.3 Sentence complexity

In broad terms, there is a direct relationship between the length of sentences and their ease of reading: longer sentences providing authors with more opportunities to construct complex sentences. The advantage of this simple measure is its transparency and, unlike the more complex ratios underlying formal readability tests (with their accompanying problems) sentence length is an attractively simple gross indicator of sentence complexity.

Whilst simpler and easily accessible sentences are clearly preferable to more complex ones, abstract writers must not allow the goal of sentence simplicity to interfere with an abstract's explanatory capabilities. A further caution on the use of this measure is that, though short sentences reflect the professional abilities of abstract writers, each abstract's clarity is also modulated by the inherent complexities of source documents and different subject areas.

The hypothesis here is that the shortest and simplest sentences are likely to be provided by the consumer oriented abstracts from the Internet subject trees, and the longest sentences will be associated with the more demanding documents from the professionally oriented sites.

Table 12. Median sentence length of the abstracts (in words)

Database	Sentence length
NetFirst	16.30 *(best)*
ADAM	16.90
Yahoo!	17.00
Infoseek	19.00
SOSIG	19.50
Lycos Top 5%	21.10
EEVL	21.30
OMNI	21.30
LISA	22.75
Magellan	22.75
WebCrawler	24.00
ISI	24.30
Excite	25.00
ERIC	25.50
Euroferret	27.00 *(worst)*

Once again, the table's figures bracket the median calculated for the online databases examined in Study 2. These had a median of 21.6 words per sentence, falling convincingly into the middle of the range of values in this table.

The hypothesis is clearly not supported in its fullest form. The (professionally oriented) Internet gateways are all among the high ranking (short sentence) databases. Scattered among the lower ranks are the abstracts from the print world, devoted to documents of greater complexity than typical WWW documents. Strangely, Internet subject trees are distributed throughout the table, making their abstracts *appear* indifferent to consumer pressures.

However, behind these results are interactions between sentence length and abstract length. The shortest abstracts often have median overall lengths (Table 10) little greater than their sentence lengths, and have only one or two sentences. Since sentences necessarily occur in discrete numbers, there is a tendency for the shorter abstracts typical of subject trees to be cramped into a single sentence.

4.1.4 Passive sentences

The strictures against writing with passive sentence structures, and perhaps the slapped wrist that so many word processors deliver for this crime, make avoidance of this feature a modern author's Grail. Abstract writers are surely no less sensitive on this point, though there is also a reasonable defence of the practice. Passive constructions are a natural symptom of reporting in the third person and the past tense, which are common in scientific, technical and medical writing, and so likely to be comfortably familiar to the targeted users.

The simplicity of this measure, its percentage scale of measurement, accessibility in word processors, and its occurrence in several of the abstracting guidelines, suggest it will be a

significant influence on abstract writers. The hypothesis is that the consumer oriented subject trees will show low scores because of their need for instant user-appeal; more professionally oriented gateways and print-originated materials are expected to show less concern.

Table 13. Average incidence of passive sentence constructions in the abstracts

Database	Passive usage
Infoseek	5.0 *(best)*
NetFirst	6.7
Lycos Top 5%	8.4
Yahoo!	13.2
ERIC	13.3
WebCrawler	13.4
ADAM	14.8
LISA	15.0
Euroferret	15.0
Magellan	16.7
Excite	17.2
SOSIG	17.6
EEVL	21.6
OMNI	23.8
ISI	40.9 *(worst)*

Average scores are used here instead of medians, because the latter is grossly influenced by the high incidence of zero scores for some databases. This causes some databases to have a median score of zero for their incidence of passive sentences, and so the table loses much of its meaning.

The Internet sources compare very favourably with the online databases of Study 2, their average level of passive sentence use among these was 37 per cent—far worse than any of the Internet subject trees or gateways.

The table's range of scores is astonishing; the worst incidence of passive sentences is eight times higher than the best. It seems that the hypothesis is being supported, but it also appears that the factor being rated in the table is the use of professional abstractors. Most of the table's best placings are occupied by databases employing professional (or paid) writers, and the worst four placings are filled by Internet gateways using volunteer writers, and by the ISI database using author abstracts. The truly extraordinary score for ISI is probably influenced by the technical nature of its source articles, as well as its policy of using only (free) author's abstracts. Surely an example of the piper playing only the tune he has been paid for? More satisfactory author abstracts are found in the ERIC and LISA databases, both of which make generous use of abstracts from this source.

4.1.5 The formal readability tests

The readability tests quantify and relate different aspects of texts, most commonly word length and sentence length, and each seems have its strengths and weaknesses, the principal features of the readability tests used in this investigation are:

4.1.5.1 Flesch Reading Ease

A test designed for lengthy items of reading material. It is intended to relate the average number of words per sentence and the average number of syllables per word in several 100-word samples taken from throughout the reading material. The version implemented by Microsoft produces test scores between 1 and 100, and what they describe as "standard writing" should have scores in the 60-70 region. The most easily read documents produce the highest scores. A typical score for the abstracts tested is about 30, corresponding to texts with a reading difficulty at about undergraduate level. [The Flesch score of this report is 34.8.]

4.1.5.2 Flesch-Kincaid Grade Level

A test relating the average number of words per sentence and the average number of syllables per word, but its scores are calculated to correspond to American school grades. Standard writing is said to have a score of seventh-to-eighth grade (equivalent to reading ages of twelve to thirteen years—which is also the typical reading level of tabloid newspapers). Again, the typical grade of the abstracts tested, at about 13.0 (say eighteen years) corresponds roughly to undergraduate level. [This report's Flesch-Kincaid grade level is said to be 12.0.]

4.1.5.3 Coleman-Liau Grade Level

Uses word length in characters and sentence length in words to generate its scores, which are calculated to correspond to American school grades. The test scores were therefore expected to reflect the Flesch-Kincaid Grade level scores. However, typical Coleman-Liau scores for the abstracts were about 16.0 (say twenty-one years)—quite markedly higher than the Flesch-Kincaid Grade Levels or the Bormuth Grade Level. As yet, no convincing reason has been found to account for this discrepancy, but it may relate to the kinds of words used in abstracts and the fact that one test measures word length in syllables and the other in characters.

4.1.5.4 Bormuth Grade Level

This test also uses word length in characters and sentence length in words to generate scores corresponding to American school grades. This test consistently produced grade levels of about 11.0 with the project's abstracts—almost as far below the Flesch-Kincaid Grade Levels as the Coleman-Liau is above them! Again, no convincing reason has been found to account for the discrepancy.

4.1.6 The origin of test score discrepancies

The Flesch-Kincaid, Coleman-Liau, and Bormuth readability tests purport to display US educational grade levels, and thus should produce scores directly comparable on a number for number basis—*but they do not meet this expectation*. The Flesch Reading Ease test is well known as an equivalent test that uses a different scale for its results, and these results should be simply transformable into grade scores for ready comparison with the other three tests— *but the study's scores are not transformable in this way*.

So why are these test scores so discrepant with their purported properties? Two sets of possibilities suggest themselves:

1. The test materials are "peculiar" in ways that breach fundamental assumptions underlying the tests' designs.

2. The implementations of the tests are faulty in their calculations or in their extraction of basic word and sentence data from the test material.

Under the first heading, it is reasonable to speculate that multiple, short, abstracts may violate the assumed conditions for the tests' original materials and purpose (general reading matter, and the measurement of student progress).

Under the second heading it is easy to speculate that tests might mis-identify paragraphs, words and acronyms, word fragments, etc. and make their calculations with invalid data. To some extent this is definitely true, with paragraph and sentence identification being the most suspect issues. (With no trace of humour, this report was digested by readability software and labelled as the proud owner of 1,719 paragraphs and 932 sentences.) However, the incidence of such problems is very difficult to estimate retrospectively.

4.1.7 What are the readability tests' most secure findings?

Readability tests may respond differently to different aspects of abstract readability (which is useful), and so scores from different tests cannot be fully reconciled (which is not useful), so the safest way to interpret the test scores is to aggregate their information in some way.

Table 14. Readability ranking of the abstracts

Flesch Reading Ease	Flesch-Kincaid Grade Level	Coleman Grade Level	Bormuth Grade Level
Lycos Top 5%	Lycos Top 5%	Lycos Top 5%	Lycos Top 5%
Infoseek	WebCrawler	Infoseek	WebCrawler
Excite	Yahoo!	OMNI	Infoseek
Yahoo!	Euroferret	Excite	Magellan
Magellan	Infoseek	ADAM	Excite
OMNI	Magellan	Yahoo!	ADAM
ADAM	ADAM	SOSIG	Yahoo!
LISA	Excite	Magellan	Euroferret
SOSIG	NetFirst	WebCrawler	ERIC
NetFirst	OMNI	NetFirst	SOSIG
WebCrawler	SOSIG	ISI	NetFirst
Euroferret	LISA	LISA	OMNI
ISI	ERIC	EEVL	ISI
ERIC	ISI	ERIC	LISA
EEVL	EEVL	Euroferret	EEVL

The effects of numerical inconsistencies are most easily removed by ignoring the niceties of individual scores but retaining the scores' rank information. Table 14 demonstrates such a simplified presentation, showing "best" results at the heads of columns and "worst" results at their feet. Ranks for each of the columns are simply the rankings of the raw test scores.

This mass of data is not easy to interpret but some basic observations will be helpful. Consistency of results is the quality most sought in this table, for it represents concurrence between four independent test methods. Whether results are consistently good or consistently bad, the fact of agreement suggests an underlying integrity.

The results for the Lycos Top 5% database demonstrate the highest possible level of concurrence: all four tests placed it first, usually by a convincing margin. Similarly, three of the tests placed EEVL at the foot of the rankings, an almost equally convincing result. What must not be construed into the table's results are concrete reasons for the rankings. At a broader level, the table displays some interesting groupings of data. Three of the tests (Flesch Reading Ease, Flesch-Kincaid, and Bormuth) place the short abstracts found at Internet subject trees (Euroferret, Excite, Infoseek, Lycos Top 5%, Magellan, WebCrawler, and Yahoo!) in their top five places. The remaining test (Coleman) places three of these subject tree abstracts in the top five. Equally revealingly, the longer abstracts associated with the databases derived from printed indexes (ISA, LISA and ERIC) are usually found in the lower rankings, and the intermediate length abstracts normal for the Internet gateways (ADAM, NetFirst, OMNI, SOSIG) occupy the middle ground. [EEVL's relationship to this pattern seems anomalous, for some reason it persistently holds the lowest rank.]

A usefully simplified summary of abstract "readability" across all the four tests can be constructed by summing the four sets of readability rankings to produce a single-number cumulated rank score for each database, to generate a table such as the one below.

Table 15. Combined readability rankings

Databases—ranked in order of readability
Lycos Top 5% *(best)*
Infoseek
Excite
Yahoo!
Magellan
WebCrawler
ADAM
OMNI
SOSIG
Euroferret
NetFirst
LISA
ERIC
ISI
EEVL *(worst)*

This can be done by assigning a rank number for each readability test result (1 for the best result, 2 for the second best, and so on), then totalling each database's four rank scores (one score for each readability test). A treatment like this has the great merit of being clear and uncomplicated for presentation: it reduces four disparate columns of results to a single orderly table.

Such a table is informative, but it can be combined with other measurements of the abstracts to make it into a powerful analytical tool. Table 16 is an example of this. It shows the transformed ranking, accompanied by the figures for abstract length (already seen in Table 10). The result is far more interesting and informative than a display of either factor on its own. It shows that the two sets of data for readability and abstract length can be made to tell a collective story, but it also shows that abstract length is not the whole of the story.

Table 16: Combined readability ranking and abstract lengths (in words)

Databases (in readability order)	Abstract length
Lycos Top 5% *(best)*	118
Infoseek	31
Excite	46
Yahoo!	17
Magellan	53
WebCrawler	45
ADAM	66
OMNI	58
SOSIG	58
Euroferret	32
NetFirst	60
LISA	92
ERIC	51
ISI	136
EEVL *(worst)*	69

In Table 16, the databases are ranked in their order of readability (as in Table 15), accompanied by their median abstract lengths. With the notable exceptions of the most readable database (Lycos Top 5%) and the longest abstracts (ISI) the story emerging from this table is of a firm correlation between high readability and shorter abstracts.

At the level of individual databases, it is probably unwise to seek detailed explanations of rankings solely in terms of abstract length, because half of the measured lengths cluster in the fifty-seventy character region. Can other associations contribute to explanations?

Table 17. Combined readability ranking and median sentence length (in words)

Databases (in readability order)	Sentence length
Lycos Top 5% *(best)*	21.10
Infoseek	19.00
Excite	25.00
Yahoo!	17.00
Magellan	22.75
WebCrawler	24.00
ADAM	16.90
OMNI	21.30
SOSIG	19.50
Euroferret	27.00
NetFirst	16.30
LISA	22.75
ERIC	25.50
ISI	24.30
EEVL *(worst)*	21.30

Any correlation between readability ranks and sentence lengths is very much more confused, probably by the interaction of sentence length and abstract length. Table 17 shows some of the longest sentences can be found among both the least readable and most readable abstracts. It would not be wise to see a correlation here (perhaps because of readability test failings in accurately measuring sentence lengths).

Table 18 introduces the final piece of information that readability tests generated for the abstracts: word lengths. Despite individual variations, abstracts in the top half of the table have words about half a character (a large amount) shorter than abstracts in the lower half. However, it is possible that the shortest abstracts (all but one from the seven Internet subject trees, are influenced by the generalised nature of their source material.

Between them, Tables 14, 16, and 18 seem to be demonstrating a clear association between readability scores, abstract lengths, and word lengths. Whether this is an artefact of the readability tests or a truly dependent relationship that users would sanction can be determined only by proper testing.

Table 18. Combined readability ranking and word length (in characters)

Databases (in readability order)	Word length
Lycos Top 5% *(best)*	4.80
Infoseek	5.10
Excite	5.00
Yahoo!	5.50
Magellan	5.20
WebCrawler	5.30
ADAM	5.60
OMNI	5.40
SOSIG	5.70
Euroferret	5.45
NetFirst	5.70
LISA	5.60
ERIC	6.05
ISI	5.60
EEVL *(worst)*	5.80

4.1.8 Is there a "best readability test" for abstracts?

This question is best answered in two parts: one related to the abstracts, and one to the tests.

As far as the abstracts are concerned, it seems that an important component of "readability," in as far it can be measured by these tests, is made up of contributions from word length and abstract length. Different readability tests can give different weightings to each of these factors, and can generate results that conflict in detail though concurring in broad outline.

As far as the individual tests are concerned, the project's experience with readability tests prompts a cautious approach to recommending a "best buy" readability test. Table 14 suggests there is little to choose between *sets of results* produced by the Flesch Reading Ease, Flesch-Kincaid, and Bormuth tests. All these tests agreed in the broad ranking of the database abstracts, though all showed some disagreements too. The Coleman test results were at variance with the other tests, and this is why it has not been commended.

4.1.9 Conclusions on the value of readability testing for assessing abstracts

1. The remarkably consistent clustering of test results into three groups (print derived, general Internet subject trees, and gateways) suggests that abstracts from these sources are conforming to guidelines recognised by their creators, though not generally published.

2. Readability scores showed a noteworthy and robust association with abstract length and word length; other associations may be revealed by more directed testing.

3. There is certainly a general relationship between the tests' readability findings and the general impressions of "good readability."

4. Word processor readability statistics are not free of suspicion, for instance a perfectly standard LISA abstract, in one paragraph, was unshakeably reported as being a text of three paragraphs!

There seems to be scant research evidence suggesting that readability scores can be linked to any of the factors that users would acknowledge as relevant to their usage of abstracts. The tests' simplistic calculation of scores places a premium on short words and sentences—which is a polite way of saying that convoluted prose filled with unfamiliar short words and abbreviations, and assembled with execrable grammar, may be able to generate exemplary readability scores. Readability scores are a mechanical assessment of text's basic properties; they are not a substitute for users' opinions. For instance, EEVL, one of the UK gateways, is consistently at the foot of the readability tables, but project workers made more favourable comments on EEVL than any other database examined!

4.2 CONTENT AND PROVENANCE OF ELECTRONIC ABSTRACTS

4.2.1 Introduction

The task of this part of the project was to delineate the accessible components of the database abstracts. In this context "accessible" meant not only points of information easily available to ordinary users but also those components that could be assessed easily by the project's student workers. Thus any matters likely to need unusual technical understanding or subject expertise were not investigated. Some examples of these points include following up instructions couched in subject specific vocabulary or downloading images from a non-WWW server.

This directed the investigation to several distinctive areas:

Coverage of source item subject content by the database abstracts

Sources of the database abstracts' content

Title

First paragraph

Last paragraph

Indexing, index term synonyms

Use or adaptation of the source item's original abstract

Recognition of enriching and constraining features in the source items

Bibliographies

Tables

Figures

Geographical constraints

Recognition of provenance and intellectual property

Inclusion of dating information in abstracts

Abstract "style" (informative, indicative, reviewing, critical)

Data was also collected for subsidiary matters, such as the incidence of structured abstracts, usage of jargon, evidence of quality-scoring of the source items, claims to originality, typography, navigation features, etc., but few of these points generated sufficient data to allow useful conclusions to be drawn.

4.2.2 Coverage of the source item's subject content

Conceptually, testing for this was a simple matter. Five "main points" were selected from a source article, and its abstract examined for mention of the points. Scoring this measure for presentation is complicated by the facts that the sample size for each database varies slightly

(because of idiosyncrasies of collection and the later removal of anomalous records as part of the data cleaning exercise), and not all source articles possessed five main points.

The scoring system adopted here is to convert the "yes" scores (meaning concepts present in the source were also found in the abstract) to a percentage, and simply to add the percentages to generate a crude number. To allow easier comparisons, the scores have been normalised to present the maximum score as 1.00 (there is no dimension to the number). When this is done it's possible to produce Table 19, showing each databases' success in discovering and representing source article concepts in its abstracts.

It could also be argued that this table shows the extent to which each database's abstractors and the project's workers concurred with each other. The counter argument is that, with one exception (at the middle of the table), each of the sets of abstracts used to compile the scores was the product of several different project workers, and this reduced the opportunity for results to be clouded by personal idiosyncrasies.

Table 19. Combined representation of source article's concepts and abstract length

Database	Representation of source article's concepts		Abstract length (in words)
ADAM	1.00	*(best)*	66
OMNI	0.79		58
ISI	0.75		136
NetFirst	0.73		60
EEVL	0.71		69
ERIC	0.71		51
LISA	0.71		92
Lycos Top 5%	0.52		118
WebCrawler	0.44		45
Euroferret	0.41		32
SOSIG	0.40		58
Magellan	0.40		53
Excite	0.28		46
Yahoo	0.26		17
Infoseek	0.21	*(worst)*	31

Data in this table are normalised with respect to the best performing database, ADAM, which project workers found had reported 64% of the concepts they identified in its source items.

Interpretation of the table needs careful thought, and a reminder that several layers on information are hidden behind the figures in the column headed *Representation of source article's concepts*. Overall, project workers found about 36 per cent of their source item concepts were present in the abstracts.

The best-supported interpretation of the table seems to be that databases with the largest abstracts are the most likely to include a source item's significant points. Additionally, the free Internet gateways and print sourced databases (all with longer abstracts) are markedly

more successful at making good use of their words, whereas the Internet subject trees report disproportionately fewer concepts with their smaller number of words.

An interesting fact not illustrated here concerns the progressive pattern of matches between points in the source and points in the abstract. Whilst it was easy for project staff to recognise an obvious "worthy" point as the first of their five points, later points became more difficult to identify. Naturally, there was a strong expectation of a reduced probability for the later points eventually being found in the abstract. This expectation was upheld—except for ERIC. For some reason, project workers were able to identify a larger number of ERIC's later points than the supposedly more obvious earlier ones. The reason for this is entirely obscure.

4.2.3 Sources of abstract content

The producer guidelines generally say "avoid repeating information found in other parts of a record," particularly the title and the indexing which users are likely to read. Only the National Library of New Zealand suggests, in positive terms, that synonyms should be used for descriptors; however, it also says abstracts should stand-alone and be comprehensible without the title or descriptors. Other guides use such phrases as "Avoid repeating ..."

Title information is the most likely to be repeated; this happened in more than half of all abstracts. Given the value of titles for retrieval, this is no surprise. In the table below, the ranking of databases assumes repeating title information is a "good thing" (from the providers and searchers points of view) as it is a general safeguard for *retrieval* needs.

Table 20. Repetition of title information in the database abstracts

Database	Repetition of title information (%)	
Lycos Top 5%	85	*(best?)*
ADAM	84	
SOSIG	79	
LISA	73	
Yahoo	73	
Infoseek	70	
EEVL	65	
NetFirst	59	
WebCrawler	58	
Magellan	57	
ERIC	47	
Excite	30	
ISI	27	
Euroferret	20	
OMNI	13	*(worst?)*

In this table, the scores show the percentage of valid responses to a question asking if title information had been repeated in the database abstract.

However, a source item's title words are part of an Internet search engine's database and will always be searched, so the retrieval benefits of repeating them cannot be very great. There is a simple counter-argument that repetition of title words may merely reflect the difficulties of writing abstracts that do not repeat such obvious subject words.

Though the table successfully shows the incidence of repeated title information, there is no plausible pattern in the distribution of repetition practices among the gateways, subject trees and print derived databases.

Information from the first paragraph is the next most commonly repeated component, but this happens with only about half the frequency of title information repetition.

Table 21. Repetition of first paragraph information in the database abstracts

Database	Repetition of first paragraph information (%)	
LISA	73	*(best?)*
Infoseek	67	
Magellan	57	
Excite	55	
EEVL	30	
SOSIG	30	
WebCrawler	24	
ERIC	20	
Euroferret	20	
ISI	20	
Yahoo	20	
NetFirst	19	
ADAM	16	
Lycos Top 5%	6	
OMNI	6	*(worst?)*

In this table, the scores are the percentage of valid responses to a question asking if some first paragraph information had been repeated in the database abstract.

The pattern in this table is a little more evident than in the previous table, but it is not a strong pattern. Print sourced abstracts are scattered, gateways have sunk lower in the table, and subject trees have risen a little higher. However, it is doubtful that any great significance can be attached to these patterns.

Repetition of other record elements is much rarer, typical repetition levels being between one and six per cent. These levels are unlikely to be significant.

4.2.4 Recording of enriching and constraining features present in the source items

The enriching and constraining features selected for this enquiry were chosen with a mind to the practicalities of conducting the research. A limited period for training, and the difficulties of ensuring common standards among workers operating unsupervised in different libraries, demanded that only the most distinctive abstract features be used and, for rapid analysis, data recording was limited to matters requiring only yes/no answers.

The source item features investigated were geographical constraints, bibliographies, tables, and figures. These attributes were seen as applicable to both print and electronic sources, though geographical constraints appeared particularly important for electronic sources, where US oriented sites greatly outweigh all others, and where the bias may be subtle and unintended but very important.

There is a dearth of firm evidence for constructing hypotheses in this area, so personal experience and received opinion played an unwelcome large rôle in this. Together, these surmises suggest that most print sources are accustomed to noting bibliographies and they sometimes record tables and figures, and that networked sources are more acquainted with the frustrations of national and cultural bias and will therefore play greater attention to recording significant geographical constraints.

A factor not accessible to the project's workers was the existence of a relationship between specific subject fields and particular kinds of enriching materials. For example, if one of OMNI's resources includes relevant images, would medical users be likely to take their presence for granted, and so not need them to be noted in an abstract?

4.2.4.1 Recording of source item geographical relevance and constraints by the abstracts

There is little information for anything worth calling hypothesis building in this matter. The sum of expectations was quite simple. Older print sourced abstracts may maintain their modest interest in geographical limitations—which are built into their sales and market structures. Newer networked databases, with their worldwide orientation, may show a higher level of awareness. The recent nationally focussed and funded Internet gateways will probably show the greatest geographical awareness.

Overall, geographical constraints were recognised in about 21 per cent of the resources examined, and it was a surprise to discover that the general extent of recording in the abstracts, at 12 per cent, was higher than for any other enriching characteristic.

Beyond this simple conclusion there is difficulty in a reaching hard conclusions. Internet resources, especially the subject gateways (because they are closest to specific users needs?), showed the greatest awareness, and print resources the least. There is some gentle concentration (three out of five) of the gateways in the upper half of the table, with ADAM being the most aware and OMNI the least aware of this characteristic, but elsewhere the picture is without a strong pattern. OMNI's position at the foot of the table may be more of a reflection of a low incidence of obvious geographical constraints in its (medical) source material than an adverse comment on its abstracting practices. It must also be remembered that print-sourced databases often use a filtering field to accommodate geographical constraints.

Table 22. Incidence of geographical constraints in the source, and their recording in the abstracts

Database	Geographical constraints found in source (%)	Geographical constraints recorded in abstract (%)
ADAM	47	28 *(best)*
Magellan	43	27
SOSIG	36	24
LISA	17	17
Euroferret	33	17
NetFirst	19	16
WebCrawler	27	12
Yahoo	27	10
EEVL	16	8
Lycos Top 5%	15	6
Excite	6	3
ISI	7	3
ERIC	10	3
Infoseek	10	3
OMNI	3	0 *(worst)*

In this (and later) tables, the two columns of data represent two factors: a characteristic's incidence in the source materials, and the extent of its recording in the database's abstracts. Both columns of data are absolute measures of incidence, e.g. some 47 per cent of ADAM's source articles had geographical constraints or relevance, and the point was noted in 28 per cent of the its abstracts. Because the second of these figures is based on only a sub-sample of the abstracts (those sources with the particular feature), caution must exercised in projecting the observed levels of recording onto their databases.

4.2.4.2 Recording of source item bibliographies by the abstracts

Bibliographies are one of the enriching features of documents traditionally recognised by abstracts, though there is no published evidence of the general extent of recognition. A comparative study of the levels of acknowledgement of bibliographies in print-sourced and wholly electronic databases was seen as a prime matter for the project to investigate. The hypotheses here were simply a formal statement of the traditional prejudices: that, because of their traditions, the abstracts in print sourced databases are more likely to acknowledge bibliographies than abstracts from wholly electronic databases.

Table 23. Incidence of bibliographies in the source, and their recording in the abstracts

Database	Source item included a bibliography (%)	Bibliography recorded in abstract (%)
ADAM	38	**34** *(best)*
OMNI	52	**25**
NetFirst	19	**16**
SOSIG	36	**12**
EEVL	23	**11**
Lycos Top 5%	21	**9**
ERIC	57	**7**
Yahoo	17	**3**
Infoseek	13	**0**
Magellan	13	**0**
Excite	27	**0**
Euroferret	30	**0**
WebCrawler	45	**0**
LISA	60	**0**
ISI	77	**0** *(worst)*

The striking conclusion must be that oral traditions are not worth the paper they are written on! Overall, about 35 per cent of the items examined had bibliographies, but only about 8 per cent of abstracts reported them.

Winner's laurels are plainly deserved by the Internet gateways, and wooden spoons by the print sourced databases that showed the highest incidence bibliographies in their sources but the lowest level of recording them. The dominating story of the table is one of bibliographic sins, and the greatest sinners are the traditional print sourced abstracts. Outstandingly best among the databases are the Internet gateways: individually and collectively they stand head and shoulders above the other databases, forming a firm block of best available practice. The appalling position of the ISI indexes (its abstracts ignore more bibliographies than any other database) is entirely predictable because of its author abstracts. However, this is unlikely to harm user's searches because ISI is the only database to so intensively demonstrate the presence of references in its print and screen displays. LISA, the librarians' database, shows an especially culpable attitude to recording references in its abstracts. If a database for the information and library world is so cavalier in treating bibliographies, what criticism can be made of others for failing to adopt better practices?

4.2.4.3 Recording of source item figures by the abstracts

Now that printing processes have made figures and illustrations more available in print sources, and HTML delivery systems have made them more accessible for networked information systems, figures and illustrations are an increasingly important tool for the author's expression. The incidence and distribution of figures in print and networked sources seems to have been entirely unknown before this research, and this was a major handicap in

formulating hypotheses about their occurrence and treatment. Naively, it was postulated that HTML source documents would make the highest use of figures, and Internet databases would be the most active in recording them.

Table 24. Incidence of figures in the source, and their recording in the abstracts

Database	Figures found in source (%)	Figures noted in abstract (%)	
ADAM	69	19	*(best)*
Lycos Top 5%	26	9	
NetFirst	27	8	
OMNI	42	8	
SOSIG	39	6	
Magellan	20	3	
ISI	40	3	
Yahoo	43	3	
Infoseek	7	0	
WebCrawler	18	0	
LISA	27	0	
Euroferret	33	0	
ERIC	37	0	
Excite	45	0	
EEVL	53	0	*(worst)*

The global figures are miserably disappointing: the overall incidence of figures (excluding logos, etc.) in the source items was 35 per cent, but only about 4 per cent of abstracts mentioned them.

As was true for the recording of bibliographies, ADAM was the star performer in noting figures in its abstracts. Surprisingly, EEVL—covering a discipline where graphical presentation of information is a valuable tool—was by far the worst performer in this measure. Perhaps the use of tables in source items (which was very high) is seen as so much the norm that users can take it for granted they will be employed by authors whenever appropriate?

The broad picture for this measure is not an orderly one. Four of the five gateways are placed in the top half of the table, and the print sourced abstracts are all in the lower half of the table, strongly supporting this element of the hypothesis. But the Internet subject trees are diverse in their behaviour, occur mostly in the worst part of the table, and little consistent can be said about them. One of them (Lycos) is one of the best performers, but another (Excite) has an execrable record.

4.2.4.4 Recording of source item tables by the abstracts

Tables do not occur as extensively in source material as bibliographies, but their presence is often welcomed by searchers as signifying an author's careful approach to the gathering, organising, and presentation of data, and easy access to sets of information.

If there is a hypothesis to formulate for this part of the investigation, it can be no more than an expectation that the database's attitudes to bibliographies will be echoed in their treatment of tables of information.

Table 25. Incidence of tables in the source, and their recording in the abstracts

Database	Sources with tables (%)	Tables recorded in abstract (%)	
ADAM	3	**13**	*(best)*
ERIC	43	**7**	
Lycos Top 5%	6	**3**	
Yahoo	10	**3**	
NetFirst	11	**3**	
ISI	40	**3**	
SOSIG	0	**0**	
WebCrawler	3	**0**	
Excite	6	**0**	
Infoseek	10	**0**	
Magellan	10	**0**	
Euroferret	13	**0**	
EEVL	15	**0**	
LISA	17	**0**	
OMNI	48	**0**	*(worst)*

Overall, about 16 per cent of the item's examined had tables (less than half the incidence of bibliographies), but only about 2 per cent of abstracts reported their presence.

ADAM is clearly the database most conscientious in recognising tables, even though their incidence in its source material is quite meagre. Regrettably, the majority of databases examined (60 per cent) were entirely comfortable in ignoring the presence of tables. Even OMNI, with 48 per cent of its source material containing tables, was content to ignore this information. LISA and ISI fared little better. There is little consistency in the finer details, and no wider story can to be extracted from the scattered placings.

On the evidence of this data the hypothesis fails—very convincingly. Tables in source documents seem to be undervalued by the databases, and the author's provision of tables is not seen as sufficient reason to note them in abstracts.

4.2.4.5 Inclusion of dating information in abstracts

The standard works on abstracting have nothing to say on recording a creation date for the abstract, and in the print world the question of dating abstracts is an irrelevance, because a

volume of abstracts is normally dated. Much the same applies to online databases derived from print sources.

However, in the Internet world the volatility of source items and users' opinions can be of a very different order and, because dating of source items is unusual, there is a greater need for dating the abstracts. This reasoning is the basis of the hypothesis. Databases of wholly electronic form will be more likely to carry a date with their abstracts, and commercial databases able to sell their added value services will also more likely to have dating information with their abstracts.

Table 26. Incidence of dated abstracts

Database	Incidence of dating (%)	
Excite	12	*(best)*
Yahoo	10	
ERIC	7	
Euroferret	7	
Infoseek	7	
EEVL	5	
ADAM	3	
Lycos Top 5%	3	
ISI	0	
LISA	0	
Magellan	0	
NetFirst	0	
OMNI	0	
SOSIG	0	
WebCrawler	0	*(worst)*

Since the overall level of acknowledgement of dating is only 4 per cent, the hypothesis cannot be adequately tested, but the data are interesting. Five of the eight databases with dated abstracts are commercial Internet services, two are non-commercial gateways, and one is based on a printed index—this may be significant, but the sample is too small and too erratic for it be treated as such. In the face of such low levels of acknowledgement, it is tempting to conclude that many database producers have either not considered this matter, or that they wish to conceal evidence of the date (and possible obsolescence) of their abstracts.

4.2.5 Abstract provenance and intellectual property

Except for two cases, the eleven sets of producer guidelines supplied to the project made no mention of the need to sign abstracts. Bowker-Saur had an instruction to note the use of author abstracts, and the National Library of New Zealand said to indicate author abstracts by the suffix (Auth). No other information was available on these points.

The lack of information influenced the formation of a hypothesis for testing. It seemed important to bear in mind the existence of the producer guidelines—implying that professionally created abstracts are believed superior to author abstracts. Other factors were the ANSI standard's suggestion that authors and professional abstractors can write equally satisfactory abstracts, provided the standard's quality guidelines are followed (but how are guidelines to be followed without the training ?). Earlier, Borko and Bernier[21] had taken the opposite view to ANSI, and advised in favour of professional abstractors.

The hypothesis formed from these contradictions was that, if staff abstractors were used, their work would be acknowledged to advertise its implied superiority, and that author's abstracts would also be acknowledged as credible sources. Furthermore, acknowledgement is most likely to occur in the commercially oriented Internet subject trees, the print derived databases, and the commercial gateway (NetFirst)—which all have good commercial reasons for wanting to demonstrate their added value points.

Table 27. Acknowledgement of abstract provenance

Database	Acknowledgement of provenance (%)	
ERIC	83	*(best)*
NetFirst	24	
Yahoo	17	
ADAM	13	
Euroferret	13	
Lycos Top 5%	12	
WebCrawler	9	
Infoseek	7	
OMNI	6	
SOSIG	6	
EEVL	5	
Excite	3	
ISI	0	
LISA	0	
Magellan	0	*(worst)*

The hypothesis is modestly supported, but by no great margin. The overall incidence of source acknowledgement was only 13 per cent, but the levels in individual databases varied from ERIC's 83 per cent to zero in ISI, LISA, and Magellan. One would have supposed that commercial interest and the desire to register the ownership of valuable intellectual property would have generated higher levels of observance in this matter.

The general picture is that commerce does seem to exert a governing influence. Most of the commercial Internet subject trees and the sole commercial gateway are found in the top half of the table (six of the eight cases), and most of the non-commercial gateways (three out of

---

[21] Harold Borko and Charles L. Bernier. *Abstracting concepts and methods*. New York: Academic Press, 1975.

four) are in the bottom half of the table. But LISA's expensive staff-written abstracts are entirely unacknowledged, as are ISI's author abstracts.

4.2.6 Style and character of abstracts

The concept of an abstract was devised for a world of paper documents, when the abstractor could be sure that the document abstracted was identical to the one seen by users. In the networked environment the document seen by the abstractor may be updated after the abstract is written. More problematically, the user's navigation of a hypertext document can ignore the parts highlighted by the abstract, and the user may experience a document quite different from the abstractor. Worse than this for the likely relevance of the abstract is the case of a gateway's abstractor, who is frequently identifying and abstracting *collections* of documents. Collections may change, and may be experienced in different ways by different users, and each collection can point onward to more documents. Truly, in the electronic world our conception of abstracts needs a richer taxonomy.

Traditionally, abstract taxonomy has only four categories (Borko and Bernier[22] [p13-20] described additional special categories) and, in principle, they are simple to apply. But, in practice, there are important discrimination difficulties in deciding where the class boundaries fall, especially when processing large numbers of abstracts in a short time! The categories are usually interpreted as:

Informative abstracts: provides important quantitative and qualitative information from the source items, and are sometimes able to act as substitutes for the sources.

Indicative abstracts: describe and characterise the source item; this type of abstract is adapted to dealing with multi-topic items.

Reviewing abstracts: summarise items and makes appropriate critical assessment of their character and content.

Critical abstracts: analyse items and show the reader whether they are significant and worth pursuing, but do not attempt to comprehensively represent the items.

Mixed abstracts: contain significant elements of two or more of the above.

The producer guidelines did not show a marked bias on the use of indicative or informative abstracts; three allowed for a mix of styles, three had a preference for informative abstracts, two of them favoured indicative abstracts, and three others related the decision to the style of the original article.

Project workers often found it difficult to make satisfactory characterisations of abstracts, and their task cannot have been made easier by the brevity of many electronic abstracts. In general many more abstracts were characterised as informative than indicative, for both conventional databases and for the Internet subject trees. Interestingly, the Internet gateways, which make explicit judgements by including sources in their databases, also made good use of a reviewing style with further judgements. This data must be treated with caution, because of both its subjective and ill-defined nature, and the extent to which databases had abstracts with mixed styles (which are particularly prone to inconsistent classification).

If abstracts in a mixed style are analysed into their components, and their data included with other stylistic data, the overall distribution of styles across all the databases is revealed.

---

[22] Harold Borko and Charles L. Bernier. *Abstracting concepts and methods*. New York: Academic Press, 1975.

Table 28. Distribution of basic abstract styles

Abstracting styles	Incidence (%)
Informative	50
Indicative	30
Reviewing	19
Critical	1

The overwhelming prevalence of informative and indicative styles is very apparent.

The data on abstract styles can also be examined from the perspective of the individual databases, but only in respect of the informative and indicative styles (there is insufficient data for a wider examination).

Table 29. Incidence of the informative style in abstracts

Database	Abstracts using an informative style (%)
Magellan	97
Lycos Top 5%	76
LISA	73
ADAM	72
Euroferret	70
Infoseek	70
ERIC	67
OMNI	65
WebCrawler	61
ISI	60
EEVL	55
Excite	45
NetFirst	35
Yahoo	33
SOSIG	30

The percentages shown here have been calculated from all the decisions made about abstract style, and have been normalised to allow for differences in sample size.

Table 30. Incidence of the indicative style in abstracts

Database	Abstracts using an indicative style (%)
OMNI	100
EEVL	77
ERIC	66
SOSIG	61
NetFirst	49
Excite	48
Yahoo	40
ISI	37
Infoseek	30
WebCrawler	24
Lycos Top 5%	15
ADAM	13
LISA	3
Euroferret	0
Magellan	0

The percentages shown here have been calculated from all the decisions made about abstract style, and have been normalised to allow for differences in sample size.

The two tables should be seen as complementary, if not reciprocal. Informative style is used more widely by all three kinds of database (print sourced, gateway, and Internet subject tree), but the indicative style is clearly favoured by the gateways—except for ADAM. Inspection of ADAM's abstracts suggests the reason for this may be the obvious one, a significant effort is made to put factual information into the abstract, rather than adopt the looser terminology of an indicative style. The preference of Internet gateways for the indicative style is most probably a consequence of the nature of their source material, which tends to be collections of documents rather than individual documents.

Overall, about 15 per cent of abstracts were of mixed style; Lycos, in particular, adopts a varied style with 67 per cent of its abstracts recorded as being of more than one style. The measuring instrument used in the survey did not attempt to discover the weighting of styles in mixed abstracts, so this type of abstract will not be examined further.

4.2.7 Abstract structures

The normal structure for abstracts is a simple single-paragraph presentation, but occasionally multi-paragraph presentation is used. However, in recent years there has been considerable interest in *structured* abstracts, for instance, Trawinski,[23] Salager-Meyer,[24] and Harbourt,

---

[23] Bogdan Trawinski. A methodology for writing problem structured abstracts. *Information Processing and Management*, *25*(6), 1989, p693-702.
[24] Françoise Salager-Meyer. Medical English abstracts: how well are they structured? *Journal of the American Society for Information Science*, *42*(7), August 1991, p528-31.

Knecht, and Humphreys.[25] When such abstracts are used, it is expected that all or most of a database's abstracts will have a similar, visible, sub-headed structure. Abstracts of this type might explicitly organise their content under sub-headings such as Problem, Case history, Methodology, Findings, etc. Structured abstracts have mostly been reported in the context of medical work. Experience suggests that, so far, the majority of printed and online database producers have preferred to retain their traditional format.

The project tried to establish the incidence of structured abstracts among the newer Internet databases. There is insufficient surrounding evidence or information to re-phrase the question in the form of a hypothesis.

Sadly, the data produced in response to this question was quite unsuited to answering it. A few of the project's workers could see structure with worryingly ease, some could see it when a pattern was repeated without the presence of headings, and others reported structured abstracts only when explicit headings were present. One medical database, HealthStar, included structured abstracts in about half of its of records, and World Textile Abstracts used some structured abstracts too. LISA abstractors sometimes seemed to follow a distinctive pattern or checklist, but the system's abstracts were not presented with formal sub-headings expressing checklist's points. Much the same is also true of other conventional databases, as revealed in the study of producer guidelines.

No database examined consistently followed a structured model. Six of the eleven producer guidelines either specified or suggested by example that a traditional single-paragraph structure should be used, and this is largely born out by observation. Single-paragraph abstracts are used exclusively in 28 (70 per cent) of the resources examined while a further five have single paragraphs in over 90 per cent of cases—making some eighty per cent of the resources that are entirely or predominantly single-paragraphed in format. (Measurements of the incidence of single-paragraph abstracts were generated by readability test statistics, but there is some uncertainty about the accuracy of the figures.)

---

[25] Anna M. Harbourt, Lou S. Knecht, and Betsy L. Humphreys. Structured abstracts in Medline, 1989-1991. *Bulletin of the Medical Libraries Association*, *83*(2), April 1995, p190-5.

5.0 THE RESEARCH QUESTIONS REVIEWED

5.1 Introduction

It is now possible to make sensible responses to the research questions raised in Section 1.3. Most were more encompassing than the studies of particular kinds of data that fill so many pages above, and answering them usually requires integration of evidence from several of these more specific studies. Because of this integration, the question's answers are normally less specific than the discussions of data, and are as brief as possible, for easier comparison. All their points are amplified in earlier discussions of the data from Study 2.

5.1.1 Are the print sourced database abstracts usefully summarised by the measures used, and are their abstracts' properties revealed?

All three print sourced databases tend to use long abstracts, long words, long sentences, and suffer from more than average levels of passive sentences. It is hardly surprising that the readability tests consistently rank their abstracts among the least readable tested. Their representation of concepts from the source items is better than average, but their acknowledgement of enriching characteristics is worse. These databases are certainly similar, and their regard seems to be based on their length and informativeness.

5.1.2 Are author written abstracts different from the professionally written abstracts used by ERIC and LISA?

It was optimistic to ask this question. Though ISI uses wholly author abstracts, both ERIC and LISA use a mixture of author's and abstractor's abstracts, and the balance of their contribution is not sufficiently closely identified for valid investigation. However, compared to the abstracts in ERIC and LISA, it is clear that ISI's "pure" author abstracts are considerably longer and more passively expressed. In other matters there is little to consistently differentiate the abstracts from these databases

5.1.3 Which properties of Internet subject gateway abstracts seem to correlate with users high opinions, and how do these properties differ from abstracts converted from the print environment?

The distinctive properties of these gateways' abstracts aren't easy to pin down. The simple physical properties of their texts are about average, except for a strong penchant for passive sentences, and the readability scores are average or worse. Where they all excel is in their sensitivity to the content of source items. They showed the highest abilities in capturing and conveying topics from source items (despite the modest average length of their abstracts), and generally displayed a useful sensitivity to the presence of bibliographies, etc. The difficult part of this question relates to possible correlations. None of the factors examined suggested a high correlation with users reputed opinions, and the possibility of such a correlation would

be impossible to prove without data gathered from users. Indeed, the opinions may be a reflection of subject coverage and database selectivity, not abstract content.

5.1.4 Are the general Internet subject trees as truly similar as they first appear, and how do they differ from longer reviews?

With the exception of Lycos Top 5%, abstracts at the general Internet subject trees do have a common identity. They are certainly variable in nature, but they pass the old classifier's test of similarity: they form a natural group whose members have more in common with each other than with those outside the group. Their broad properties can be summarised as: brevity (about 40 words in length), composed with short words in long and passive sentences, attaining the highest levels of readability, but failing to discover and represent source item concepts or enriching features. The most important correlation within this envelope of properties must be between these abstracts' shorter than average length and their lower than average ability to convey the subjects and features of source documents.

These points deserve the obvious rider: simplicity and brevity may be worthwhile qualities of abstracts able to favourably influence users' initial opinions, but they almost certainly constrain retrieval effectiveness.

5.1.5 In what ways do the abstracts at Lycos Top 5% differ from those at other general Internet subject trees?

In most respects, abstracts from the Lycos Top 5% database have the advantages and disadvantages of any Internet subject trees. What distinguishes the Lycos abstracts is the frequency with which they top the "score charts" of features likely to influence the overall effectiveness of databases. Compared to other subject tree abstracts those at Lycos are: much longer (more than twice as long), use shorter words and fewer passive sentences, are notably more readable, cover more of their source item's concepts, and are more likely to mention title information, bibliographies, figures, and tables. No one of these features is invariably diagnostic of a better abstract, but collectively they form a powerful specification and go a long way to explaining Lycos' long held reputation among subject trees.

5.1.6 Does the brevity of Yahoo's abstracts markedly affect any of its properties?

As well as having the shortest abstracts and sentences of any subject tree, Yahoo has the longest words, is somewhat better than an average subject tree at identifying bibliographies, tables, and figures, but is almost the worst subject tree at capturing source item concepts (unsurprising, given its minimal abstract length). There is little in the project's data to account for Yahoo's position of high regard—which must originate in factors such as size of database and ease of navigation. If the generous length of its abstracts was significantly responsible for Lycos' high regard, then the very short length of abstracts in Yahoo's highly reputable database must imply that either their reputations or the rôle of abstracts has been misunderstood or inflated.

5.1.7 Have the economic pressures on Euroferret affected the measured properties of its abstracts?

For the researcher, this database can be seen as an "in progress" experiment. Starting from a very disadvantageous position, this database invited its users to submit both abstracts and recommendations for source sites. So the researcher can see a database in the process of

creation, without benefit of experience or significant abstractor guidelines (though abstracts are probably edited). The result is visible in its files. For a subject tree, its abstracts are among the briefest and its words among the longest, its sentences are long, passive and among the least readable, and it weak at recording tables, figures and bibliographies. But despite these weaknesses Euroferret is surprisingly average at finding and identifying subject concepts in source items.

5.1.8 In what ways are abstracts from print sources, gateways, and subject trees differentiated from each other?

A large number of factors are summarised in the table below. Only a very willing observer would claim to see a strong pattern of causes and effects in this distribution of assessments. But it is worth noting some associations:

- The print tradition's longer abstracts, sentences and words have probably been the main causes of its poor readability and its good ability to capture source item concepts.

- Internet subject trees probably owe their good readability and poor capability of capturing source item concepts to precisely the opposite properties.

- Internet gateways have developed an excellent overall position by avoiding the extreme practices shown by print-based sources and Internet subject trees.

- Internet gateways show the greatest overall sensitivity to the content of source items.

Table 31. Summary of the properties of electronic abstracts from different sources

Properties of abstracts	Print-sourced online databases	Internet gateways	Internet subject trees
Abstract length	**LONGEST**	Medium	*Shortest*
Word length	Long	Long	**SHORTEST**
Sentence length	Long	**SHORT**	Varies
Passive sentence usage	Varies	*Worst*	Varies
Readability scores	*Worst*	Medium	**BEST**
Recording of a source item's concepts in its abstract	**BEST**	**BEST**	*Worst*
Recording of a source item's bibliography in its abstract	*Worst*	**BEST**	Average
Recording of a source item's tables in its abstract	Varies	Varies	Average
Recording of a source item's figures in its abstract	Varies	**BEST**	Varies
Recording of a source item's geographical relevance or constraints	Varies	Varies	Varies

For simplicity of presentation the table has "summaries of worth" for each feature, with the most valued assessments in **BOLD** capitals and the least valued in *italic*.

6.0 CONCLUSIONS

6.1 Conventional online databases and print-sourced abstracting services

The survey of producer's guidelines showed very clearly that author abstracts are used freely by major players such as Elsevier and ISI, and this is paralleled by other producers, who advocate or support the use of author abstracts or slightly-amended author abstracts whenever they are suitable. However, this is not a universal approach, for example, the Royal Society of Chemistry note that "the abstract required for an RSC product will seldom be identical with that found in the original article." The use of author abstracts results in their intrusion into this research, and it cannot be ignored. For instance, this may be an important reason behind the poor scores of LISA, ERIC, and ISI databases in readability rankings in Tables 14 and 15. Author abstracts are used exclusively by ISI, and are sometimes used by LISA. ERIC's practice in this respect is not known, as guidelines did not arrive in time, but barely amended author abstracts were noted.

While the results would generally seem to show the successful transition of a paper-based tool into electronic databases, it is interesting to note producer's general reluctance to date their abstracts, despite this being a significant and obvious new requirement for the world of electronic information.

6.2 Character of abstracts

Abstracts for the Internet subject trees and gateways are not simple extensions of the older print derived abstracts seen in databases such as ERIC and LISA. On average, the new resources have abstracts about 50 words long, and their sentences and words are shorter than those in the print sourced online databases. Identifying and describing the most significant aspects of these differences will require more detailed investigation involving real users needs and reactions.

It is also worth noting that the explosive growth of electronic resources means that there are many more users of such resources, and that a majority of them are not information workers or skilled researchers; in fact, such is the success of the Internet that many users may be children. There has been a considerable amount written about the searching of databases by intermediaries and end users,[26] initially for online databases and later for CD-ROMs, and there is little need to repeat it here. What may be stated here is that existing abstracts pay very little heed to unskilled or naive users' problems. These users rarely have the evaluative skills needed to enable them to sort good resources from poor resources. For these people, it is increasingly the task of abstracts—particularly those at Internet gateways and subject trees—to guide them in resource discovery.

There is also a philosophical distinction to be made between abstracts such as those found in ERIC or LISA and the new abstracts found in Internet subject trees and gateways. While the former largely fulfil a filtering (out) rôle in which users scan through the abstracts in a

---

[26] Plutchak, T. Scott. On the satisfied and inept end user. *Medical Reference Services Quarterly*, *8*(1), Spring 1989, p45-8.

broadly appropriate answer set in order to reduce it to the ideal records, the latter are more active in the area of information discovery or filtering *in*. This resource classification or descriptor role may to some extent be subsumed by use of metadata (Dublin Core and Warwick Framework) but, at this stage, it seems that the distinction is useful and may play some part in defining the variations between the two broad groups. Because of their origins and purpose, the eLib products (OMNI, SOSIG, ADAM and EEVL) probably fall midway between the two extremes. That is to say, because of their highly selective "acquisition" of resources much of the filtering *out* is pre-coordinate and its purpose is to bring together (or filter *in*) highly subject-relevant resources. Confronted with relatively small answer sets, users may be likely to use the abstracts to both filter *in* and filter *out*.

6.3 The readability results

In the research design, considerable weight was attached to the usefulness of word processor readability testing capabilities. This seemed to offer an obvious and accessible pathway for analysis of the abstracts. However, results were not as clear, consistent, and useful as was hoped. It isn't clear where the reasons for this disappointment are to be found. Some suggestions are:

- Conversion of the displayed HTML abstracts into "clean" text for testing may not always have been perfect, and the converted text may not have been perfectly adjusted to the word processor's expectations (these are not detailed in the software's help files).

- The word processor's implementation of each test's algorithms may have been imperfect.

- The readability tests, which are heavily dependent on the ratios of quantities such as word and sentence lengths, numbers of syllables per word and words per sentence, may not respond reliably to text samples as short and variable as typical abstracts.

As discussed above, the project was not the first to make this discovery. Readability test scores are very enticing for their ability to convert complex matters into single-number scores, and useful for their by-products of basic counts of words, sentences, etc. but they not infallible research tools, and some basic research is needed to determine their value for abstracts and other information materials.

6.4 Internet resources

There is a further and related distinction between the use of abstracts in conventional databases and in Internet subject trees and gateways. Bibliographic databases usually contain fielded data—the information they contain is physically split into elements such as Title, Author, Publication Date, or Abstract—while the HTML sources in AltaVista, Yahoo, Excite, etc., do not have "fields" that enable searchers to restrict their search to a particular element of data. This generalisation is only broadly true: some Internet resources may offer access to databases that *apparently* contain fielded information by building this element into their link structures. EEVL allows broad filtering by, for example, the resource-type field, so does OCLC's NetFirst, but the other Internet gateways do not allow such filtering.

This is important as it leads to the consideration of the purpose, parameters and limits of an abstract for the Internet. In this medium, if not in conventional databases, there may be a need to move beyond the established ANSI model of an abstract. An ideal "Internet abstract" might include, for example, user guidance, assessment of authority, discussion of physical

attributes (the design of the site or the ease of navigation), judgements on quality, pointers to alternative sources. It is an open question as to whether extension this is appropriate for an abstract; the additional complexity may outweigh any advantages.

An alternative approach to these problems may be possible within the area of metadata, though at this time and without appropriate standards it may be impracticable. Most Web users are familiar with resources that try to cheat the search engines by excessive use of repeated synonyms in the metadata area.[27] In a similar but more constructive manner, abstract-like functions and filtering information could be attached to resources as metadata. The attraction of this approach is in the possibility that other metadata elements could establish the provenance and date of the abstract while, as has been suggested by the Centre for Information Quality Management (CIQM), still others could contain qualitative and evaluative information.[28]

6.5 The future

6.5.1 The user's view of abstract quality

While this project's work has established an important series of abstract parameters, they are essentially a series of snapshots focused on a small set of resources as they exist in the summer of 1997. Whatever their mission statement, or their variety of content and style, it cannot be denied that existing abstracts are an apparently successful means to an end. The word "apparently" is, however, important, and we believe that a next logical stage would be an investigation to examine **abstract success**.

This study has not taken account of users reactions or valuations. Questions such as "What makes an abstract successful?" and "What would users like to see in an abstract?" have to be asked. These can be extended to look at the relative merits of abstracts organised as plain text, multi-paragraphed, or formally structured, and the user view of abstracts which move beyond the ANSI definition to "ANSIplus" and include some of the elements suggested above. Without consulting real users and studying their opinions and behaviour, abstract research is in danger of becoming a patronising and unvalidated activity. Ultimately, only the user can say what is needed for a traditional abstract and what is needed for abstracts provided by Internet subject trees and gateways.

6.5.2 The value of readability testing for abstracts

Much data is now available from the study, and much more analysis of this is possible. But it would also be valuable to extend these data by using readability measurements calculated by other means, i.e. manually and with other packages, to verify the accuracy of the tools used by the project. Potentially, word-processor based readability tests are a valuable tool for producing more "readable" abstracts. This work should be taken further, and the wider value of these basic tests should be more completely explored.

6.5.3 Metadata and abstracts

The CIQM proposal to include metadata measuring resource quality should be re-examined and consideration given to the possibility **of metadata abstracts and abstract qualifiers** in

---

[27] Tracey Stanley. Search Engines Corner - Keyword spamming: cheat your way to the top. *Ariadne*, Issue 10. http://www.ariadne.ac.uk/issue10/search-engines/

[28] C. J. Armstrong. Metadata, PICS and quality. *Ariadne*, Issue 9. URL http://www.ariadne.ac.uk/issue9/PICS/

addition to the evaluative elements suggested. Much of the metadata work to date has centred on resource discovery but it is apparent that resource discovery and resource filtering are two sides of the same coin.[29] Long term development in this area is being considered by W3C, who should be approached in order to co-ordinate work in this area.

6.5.4 An abstracting toolkit

A final area for future work—clearly pointed to by both the producer guidelines and the research's demonstration of tremendous variations of approach, quality and style in current abstracting practice—is the development of a **toolkit for the production of abstracts**. It should be noted that some of the "performance" variations reported here could have arisen from differences related to subject content, and not the more objective matters of morphology addressed by the project. A toolkit would need to show abstractors when and how to make provision for subject related differences. Arguably, however, a toolkit would be most relevant to the production of abstracts by voluntary abstractors, and could be adapted to provide guidance for naïve users who are unsure what might remain unmentioned because it is taken for granted.

The purpose of a toolkit is best summarised as the provision of the portable skills and knowledge needed by new abstractors and authors to more easily assess sources, and to ensure users get the key features they need to aid their evaluation of abstracts and sources.

---

[29] Paul Resnick. Filtering information on the Internet. *Scientific American*. March 1997. http://www.sciam.com/0397issue/0397resnick.html

7.0 RECOMMENDATIONS

7.1 RECOMMENDATIONS FOR RESEARCH

The project was successful in measuring a number of morphological aspects of abstracts and relating them to the origins of abstracts, but it has raised questions that can only be answered by further studies involving real or surrogate users.

1. There is a clear need to examine relationships between the measurable properties of abstracts and the value and performance of abstracts for users, so as to guide producers in the most cost-effective ways of providing abstracts of the greatest benefit for users. Such investigation will need to be partially laboratory based, to avoid the problems of uncontrolled variations in test materials.

2. Considerable variations in the properties of abstracts from different traditions and sources seem unrelated to likely performance differences, and suggest that producers are not making good use of all existing knowledge. Similarly, unexplained variations in abstract producers' guidelines show there is a need for a toolkit to guide abstract writers in the "best practices" of their craft. Such a toolkit would need to reflect the needs of the three traditions of abstract writing: print-based, Internet gateways, and Internet subject trees.

3. Particularly among the Internet subject gateways, the great variability in the recording of a source item's enriching features suggests there may be subject-related differences in abstract content, production, and use. These differences are without clear pattern, and they may be importantly affecting the use and value of abstracts. Abstract producers should consider these differences and their relationship to users needs.

4. Simple computer-based readability testing produced results of intriguingly mixed consistency. The value of readability testing as a quality control technique for abstract writers has certainly been indicated, but its practicality and the limits of its application remain an open question. This question should be examined to determine the relationships between the test variables and abstract users' perceptions of readability. The likely beneficiaries are database producers, who could gain a quality control tool, and information researchers, who could gain a better understanding of an accessible methodology for analysing text.

5. Abstracts are naturally found attached to source items as well as within secondary sources, and at present there is no formal mechanism for identifying the origin of an abstract or for distinguishing it from the body of the resource. Developments in metadata, most notably surrounding the Dublin Core, Warwick Framework, Resource Description Framework (formerly PICS—endorsed by the EU in their October 1997 report[30]) and the London Quality Framework (LQF) proposed at a recent CIQM/LA meeting, may be the most satisfactory way to achieve this recognition. The abstract and abstract qualifiers should be considered in these discussions, and representations made to relevant parties.

---

[30] Reported in *BNA Daily Report for Executives*, 17th October, 1997, A4. See also: *Illegal and harmful content on the Internet: communication to the Economic and Social Committee and the Committee of the Regions*. http://www.echo.lu/legal/en/internet/content/communic.html

The authors recommend that MODELS brief UKOLN, W3C, and LQF representatives accordingly, and maintain a watching brief over work in this area.

7.2 RECOMMENDATIONS FOR ABSTRACT WRITERS

1. Abstracts need not be lengthy to secure a valuable degree of representation of source items. The literature of abstract writing—which is heavily influenced by printed index needs—has suggested 100 or more words are needed, but Internet gateway abstracts averaged about 62 words (barely half that of online databases) yet scored as well or better than longer abstracts for most test criteria.

2. The three distinctive schools of abstract writing (print sources, gateways, and subject trees) have their established practices, but these appear founded on habit not performance requirements. There is scope for all schools to learn from each other.

3. Readability tests provided in word processors can be frustratingly fallible, but can also be a valuable quality control tool when used sensibly.

4. Automatically generated readability advice on passive sentence usage may not be appropriate for all subject areas. Some of the most successful abstracts (from Internet gateways) "suffered" from high levels of passive sentences.

5. Informal comments from project workers strongly suggested that an abstract's information content was one of the most influential key features distinguishing useful from non-useful abstracts.

8.0 APPENDIX: READABILITY TEST SCORES

8.1 Raw readability scores

Median scores are presented here to emphasise the variability of the readability data, even when smoothed by presentation of medians rather than raw scores for individual abstracts. Any revelation of database abstract variation that might be read into this table is entirely unintentional, and should not be pursued with excessive enthusiasm.

Table 32. Median readability scores

Database	Flesch Reading Ease	Flesch-Kincaid Grade Level	Coleman Grade Level	Bormuth Grade Level
ADAM	31.50	12.45	15.30	10.90
EEVL	13.90	15.50	18.80	11.50
ERIC	18.55	14.40	19.35	11.00
Euroferret	25.70	12.00	21.85	10.95
Excite	41.70	12.60	14.50	10.90
Infoseek	41.90	12.10	13.80	10.85
ISI	23.90	15.30	17.95	11.40
LISA	28.85	13.90	18.35	11.50
Lycos Top 5%	52.65	9.90	12.40	10.45
Magellan	33.85	12.25	16.40	10.85
NetFirst	28.30	13.00	16.60	11.10
OMNI	32.60	13.20	14.40	11.20
SOSIG	28.60	13.80	16.00	11.10
WebCrawler	27.70	11.90	16.60	10.70
Yahoo!	36.60	11.95	15.75	10.90

Median values are presented for readability scores, since these have fewer statistical problems when summarising numbers, which are already ratios.

8.2 Interpreting the raw scores

It is normally taken to be a statistically improper act to further process the results of earlier statistical processing, such as by taking averages of averages, and it clearly shouldn't be done. Nevertheless … the temptation is great, surrender is easy, and the results can be very interesting.

The small table below shows the averages of the median scores seen in Table 15. It also shows the standard deviation of the medians, and the ratios of the standard deviation to the averaged score for each test's results.

Table 33. Averaged readability scores and their standard deviations

	Flesch Reading Ease	**Flesch-Kincaid Grade Level**	**Coleman Grade Level**	**Bormuth Grade Level**
Averaged median scores	31.09	12.95	16.54	11.02
Standard deviation of median scores	9.63	1.46	2.43	0.29
Ratios of standard deviations and averaged scores	31%	11%	15%	3%

The table makes it very clear that caveats must be placed on any direct interpretation of the median readability scores. The Flesch Reading Ease scores and the Flesch-Kincaid Grade Levels both generated test scores that, intuitively, seem to be appropriate. But the Flesch score's standard deviation is very high, and this must throw doubts on the reproducibility of its scores with a different sample of abstracts, and perhaps on the correctness of the sample size used for this test (the abstract gatherers' targets were about 30 abstracts per database). The Coleman-Liau and Bormuth Grade Levels used somewhat different data for their calculations, but produced contradictory US grade scores (these can be converted to age-related scores by adding five (years). Though the Bormuth test generated very consistent scores, with a remarkably low standard deviation, its monolithic consistency raises a question as to its sensitivity for this kind of test material.

The presentation of readability rankings in Table 14 probably has a more useful view of the readability scores. In that table, the Coleman test is the one that deviates furthest from the rude norm of the other three tests, and its value is seen as the most open to question. The other three tests are not in perfect harmony, but there is little to choose between them on the basis of the project's evidence.