A METHODOLOGY FOR WRITING PROBLEM STRUCTURED ABSTRACTS

BOGDAN TRAWIŃSKI
Technical University of Wrocław, Department of Information Systems,
Wybrzeże St. Wyspianskiego 27, 50-370, Wrocław, Poland

(Received 22 April 1988; accepted in final form 24 February 1989)

Abstract — Document problem: How to write the abstract of a scientific document so that it can better fulfill its functions in the process of information transfer? Problem solution: This method is based on a psychological model of a problem solving process. The abstract is structured and reflects main phases of this process (i.e., the information layer of a document) as well as the manner and scope in which this process was presented in a document (i.e., the formal and structural layer of a document). The problem structured abstract (PS abstract) consists of five separate parts: Document problem, Problem solution, Testing method, Related problems and Content elements. It was shown that PS abstracts as compared with INSPEC abstracts carry more information about problem solving process. Testing method: 80 PS abstracts of scientific papers were prepared and compared with the abstracts of the same documents included in INSPEC database with respect to content and length of abstracts. Content elements: ORI THY LZM HYP IDS SOL ILU JUS IDT CRI OBJ TAB EXP GEN USP.

1. INTRODUCTION

Although the history of abstracting goes back to ancient times [1] and nowadays using abstracts has become a common practice and even a necessity, the quality and usefulness of abstracts are still not satisfactory. At present, almost every abstracting service or scientific publisher has developed its own principles of writing abstracts. Borko and Bernier [2] mentioned that these principles are designed for specific disciplines and determinate groups of users. Borko and Chatman [3] having surveyed instructions for abstractors issued by 130 various scientific publications, concluded that an adequate abstract of a research article should cover purpose, method, results, conclusions and specialized content. Standards for writing abstracts based on this study (e.g., ANSI [4] and ISO [5]), contain analogous recommendations. However, the investigations carried out by Buxton and Meadows [6], Endres-Niggermeyer [7] and Milas-Bracović [8] concluded that abstracting rules used by various institutions are inconsistent and differ in their precision and adequacy. The rules are not always followed in practice. The authors also found that the descriptions of research methods are included into abstracts statistically with less frequency than they occur in full-texts of documents.

Informative and indicative abstracts are the most popular types of abstracts. Abstracts having both indicative and informative components are also often used. In the latter, the less important aspects, from a given point of view, are treated indicatively and more fundamental aspects—informatively. Rowley and Turner [9] share the opinion that using these abstracts can ensure maximum transmission of information with the greatest brevity. Nevertheless, in the texts of indicative-informative abstracts one can find trivial statements concerning facts known undoubtedly by the majority of users. For example, the abstract issued in INSPEC database ([10], abstract no. 18530) contains this statement:

A large part of the information supply to scientists and technicians in research institutions is obtained from publications in scientific journals.
Moreover there are also statements such as ([10], abstract no. 18677):

A numerical example is presented which demonstrates the validity of the theoretical results.

Statements of this kind are relatively lengthy, but do not carry too much information.

Endres-Niggermeyer [7] states that knowledge about abstracts is still incomplete and that the abstracting process has not been entirely understood yet, which has an impact on the usefulness of principles in writing abstracts used in practice. Thus, the following questions are still valid: “What should be described in abstracts of documents and how, so that they better fulfill their functions in the process of information transfer?”

2. FUNDAMENTALS OF THE METHODOLOGY

Problem structured abstracts (PS abstracts) are designed primarily for researchers, i.e., for those users who solve scientific problems. Abstracts constitute intermediate links between documents and users in the process of communicating scientific information by means of documents. Having read the abstract of a document, a scientist makes a decision as to whether the document is going to be useful for him/her at all, i.e., whether he/she should read the full text of the document or whether the information contained in the abstract is sufficient. Scientist's information needs may be multifarious and very specific, and satisfying them depends on both the document content and the manner in which this content is presented. The scientist seeks information produced in the process of scientific research and receives it in a written form through a document as a carrier.

It is obvious that not all ideas, data, results, conclusions etc., considered or obtained by the author are presented in the scientific document. On the one hand, authors select information they want to communicate using appropriate means of its expression and directing the document content to a given group of users. On the other hand, editors of scientific publications tend to limit the volume of articles and approve for publication the material judged as most important according to principles and criteria established by them. Papers presenting the results of finished investigations are preferred. Moreover, as Ziman [11] emphasizes, the author of a document presents the course of investigations and the results he/she obtained in such a way that they might be accepted by other scientists or experts in a given field. Therefore he/she does not inform of false starts, mistakes, unnecessary complications, difficulties and hesitations. On the contrary, his/her procedure is shown as simple, precise, profitable and the conclusions derived—inevitable. Thus, the information transferred in the document may significantly differ from the information actually produced during scientific research. First, its quantity is considerably lower and second, it is distorted to a certain extent.

A method of writing abstracts must then be based on the definite viewpoint on a document and the characteristics of the document treated as an information carrier must also be considered. In this paper the following point of view on the scientific document is held: Two different layers may be distinguished in the document structure, namely, an information layer and a formal and structural layer. The information layer is constituted by the information produced in the process of problem solving. The formal and structural layer characterizes to an extent the process of problem solving described in the document. Both layers must be reflected in the abstract of the document and each of them should be described separately.

Rules for identifying the information layer of a document and for representing it in the abstract should be based on an adequate model of scientific research. The analysis of that model should give an opportunity to recognize the characteristics of the information necessary for creative work as well as the characteristics of the information produced during scientific research.

In the study described below, a psychological model of problem solving was used. Psychologists such as Johnson [12], Merrifield et al. [13], and Rudniański [14] distinguish several functionally different phases in the process of problem solving, and the major ones
are: Recognizing the problem, analyzing the problem situation, generating ideas of solution, verifying ideas of solution. The process has a cyclic character e.g., most frequently the failure in a given phase forces the scientist to return to previous phases.

In order to succeed in a particular phase the scientist should make a different effort, possess determinate intellectual qualities, and utilize various kinds of information. Frequently, he needs specific information which must be selected from a great amount of information on a specialized topic and then ordered adequately. Zlocevski et al. [15] believes that in each phase a scientist treats the information differently, processes different amount of information and uses it for different purposes.

Information needs for those who conduct scientific research may concern one phase, a few phases or the whole process of problem solving described in a document. A scientist, at a given stage of scientific research, will not always be interested in the full text of a document but only in a certain part of that text. I assume that the structure of the problem solving process should be reflected in the structure of the abstract.

In order to identify the formal and structural layer of a document, full texts of 80 articles on information science were analyzed [16]. During the analysis, a temporary outline of the structure of a document content was used. This structure was based on recommendations for writing scientific papers included in 16 manuals and instructions for authors and editors of scientific literature. As a result of the analysis, redundant content elements were removed and new elements, not identified previously, were incorporated into the structure. The final list of the content elements has been shown in the appendix. Frequency figures of each content element in the set of 80 articles have also been given. Which elements will occur in a given document depends primarily on the kind and the purpose of the document, the type and the complexity of the problem described, the scope of research completed, and the thoroughness of its description. It was concluded that the content elements could serve as a good reflection of the formal and structural layer of a document in an abstract. However, due to the necessity in obtaining abstracts of a sensible length, it is possible to include document content elements in the PS abstract only in the shortened form, e.g., using three-letter codes.

3. THE STRUCTURE AND RULES OF WRITING PS ABSTRACTS

The PS abstract does not have the form of a continuous text but has a modular structure—it consists of 5 separate parts: document problem, problem solution, testing method, related problems, and content elements. The structure of the PS abstract results from the considerations carried out in the previous paragraph. The first four parts describe the main phases of the problem solving process which is presented in a document. The document problem part reflects the problem recognition phase of the psychological model of problem solving, the problem solution part represents the phase of generating ideas of solution, the testing method part describes the phase of verifying ideas of solution, and the related problems part reflects the phase of analyzing the problem situation. These parts reflect the information layer of the document. In the fifth part of the PS abstract all content elements occurring in the primary document are specified. Thus, the content elements part represents the formal and structural layer of the document.

The main goal of dividing the PS abstract into five parts is:

1. To enable a scientist to concentrate on those aspects of document content which are the most interesting for him/her at a given stage of scientific research.
2. To obtain highly informative and relatively short abstracts, to give a user a transparent representation of the content of a primary document.

3.1 Document problem

This part of the PS abstract includes the formulation of the problem which has been solved by the author and whose solution was presented in the document. Since the whole content of the document depends on the problem solved and it is the problem that in the first place attracts the user's attention, the formulation of the problem must be clear and
precise. Therefore, it is assumed that problems should be expressed in terms of interrogative statements—questions. Several authors, e.g., Bunge [17], Cackowski [18], Rozwadowski [19], and Tondl [20], have proposed to formulate problems by means of questions. Basic notions considering the formal structure of questions used in the present paper have been taken from Ajdukiewicz [21]. When representing the problem by means of a question it is essential to determine precisely the question marker and formulate clearly the datum quaestionis. For example in the PS abstract of the document [22] the following document problem was formulated:

What factors determine response time variations in an online information retrieval system?

The formulation of a document problem may contain information characterizing the problem situation more precisely, e.g., known data, assumptions or limitations. In this case a conditional question may be used. For example in the PS abstract of the article [23] the document problem was formulated as:

Given that the term significance weights have been estimated using within-document frequencies, does the incorporation of the weights in the document retrieval system based on a probabilistic model result in improving search performance?

Feature, properties or purposes of the problem solution may be closely characterized in a question by using appropriate phrases or clauses preceded by “to,” “in order to,” “so that,” “in order that.” For example, the PS abstract of the paper [24] contains the following formulation of the document problem:

How to represent structural formulas in a chemical information retrieval system in order that searching a specific substance or a family of substances, of which query compound is a member, should be possible?

It is allowable to formulate more than one problem in the document problem part. One may distinguish two basic kinds of relations between problems which solutions are described in the document, i.e., subordination and coordination. The former is the relation of the type “main problem-subproblem.” In regard to this relation the following groups of subproblems can be found:

1. Subproblems formulated in order to examine the assumptions of the main problem.
2. Subproblems formulated in order to recognize and analyze the problem situation which is represented by the main problem.
3. Subproblems formulated in consequence of the decomposition of the main problem into partial problems. The solutions of partial problems constitute the solution of the main problem or are used to achieve this solution.

Similar ideas were developed by Rozwadowski [19].

3.2 Problem solution

This part of the PS abstract includes a statement(s) which constitutes the proper answer to the question representing the document problem. The conformity of the question placed in the document problem part to its proper answer described in the document and written in the problem solution part constitute the fundamental criterion of judgment whether the document problem and the problem solution are formulated correctly. It is evident that this criterion is not sufficient. Further criteria for a well-formulated problem are given by Bunge [17] and Cackowski [18]. For example in the PS abstract of the article,
"Determining the Required Number of Online Catalog Terminals: A Research Study," [25], the following document problem and its solution have been given:

**Document problem:**

How to estimate the optimal number of online catalog terminals?

**Problem solution:**

The method consists in determining whether a catalog use pattern, established on the basis of data from an existing system, matches any of the standard statistical distributions (such as the Pearson Type III and the lognormal distribution). Then the number of terminals needed under a variety of conditions is calculated using queuing models.

The problem solution may be better characterized by giving additional information included in such content elements of a document as "justification of solution," "characteristics of solution," "presentation of calculation results," "description of negative results obtained," "comparison with results obtained by other authors," "possible usage areas in practice," "possible usage areas in science" (see appendix).

Obviously, it will not be possible in every case to include an exhaustive proper answer to the problem document in the *problem solution* part. For instance, when the problem of one of the types: "how?," "what?," and "why?" is formulated, the proper answers to such questions are usually expressed in the form of many statements or phrases. An attempt to include all of them in the *problem solution* part will result in exceeding the reasonable or established limits of the PS abstract length. One may then describe only theories or principles on which the problem solution is grounded or give basic characteristics of the solution. For example, in the PS abstract of the paper "A Model of Feedback in a Document Information Retrieval System" [26], the following document problem and problem solution were formulated:

**Document problem:**

Given that in an information retrieval system documents are represented using vectors of weighted descriptors, how to realize automatically a relevance feedback in the system?

**Problem solution:**

The algorithm uses the theoretical apparatus of fuzzy sets for discriminating between semantically uniform subsets of documents amongst those retrieved for a given query.

### 3.3 Testing method

The testing method should be characterized by describing the idea of the method, the scope of testing and the evaluation criteria used. For example, in the PS abstract of the article [27] the testing method was presented as follows:

Time, cost, recall and precision of the searches performed on 20 topics from the biomedicine field by 16 users using the CONIT automatic intermediary system and with the help of human expert intermediaries have been compared.

In the above example, the idea of testing method is the comparison of (specified features of) searches on specific topics performed by users using an intermediary system and by human expert intermediaries. The scope of testing is determined by giving numbers e.g., 16 (users), 20 (topics) and names, e.g., CONIT (an intermediary system), biomedicine (a subject-matter field). The evaluation criteria are time, cost, recall and precision.

The approach outlined above refers to those phases of verifying solution ideas in the course of which the experiments, observations, measurements, etc. were conducted in order to obtain necessary data.
3.4 Related problems

This part of the PS abstract should include the problems which have been admittedly formulated in the text of a document but solutions have not been found and presented there. Such problems may be:

1. Problems which were decomposed into sub-problems. Only the solutions of the sub-problems were presented in a document. In this case the formulations of sub-problems should be placed in the document problem part and the main problem—in the related problems part.

2. Problems which, in the document authors' opinion could be solved using the solution of the document problem. These problems are usually defined within such content elements as “possible usage areas in science,” “possible ways of improving solution.”


For example, in the PS abstract of the document [28] the following problems have been formulated:

**Document Problem:**

How to classify fiction, so that it may be retrieved according to user's value criteria?

**Related problems:**

What efficiency measure is adequate to evaluate fiction retrieval?

3.5 Content elements

This part of the PS abstract represents the formal and structural layer of the document. All content elements occurring in the document that are abstracted should be enumerated in it. To facilitate the analysis of the document, a list of content elements was prepared. The methodology of arriving at this list was shown in the previous paragraph. One needs only to identify which elements from the list occur in the content of the document.

The description of the phases of the problem solving process does not give a complete representation of the document. The content elements of the document text serve to complement this representation by offering a view of the process of problem solving as reflected in the primary document. The user's information needs may be multifarious and peculiar. He/she might be interested only in such documents in which apparatus or materials used by testing were described or in which the analysis of measurement errors was carried out. On the other hand, information about documents where the origin of the document problem was explained or about theories constituting the context of the problem or about the history of previous research might be decisive for selecting such and such a document, particularly by the user who does not possess enough knowledge in a given field. One user would like to obtain a document in which solutions are presented in the form of algorithms and another one might prefer documents with problems and their solutions described mathematically. Due to the content elements part, the PS abstract enable any user to select documents according to his/her individual needs and preferences. To achieve this aim, one does not need to place the content element names in the PS abstract in their full lengths. I am suggesting that content elements could be denoted by means of three-letter codes in the PS abstract. The list of content elements with codes assigned to them has been shown in the appendix. The use of short codes will increase the abstract length to a small extent only. Each code is unique and corresponds to one content element only. The codes are mnemonic abbreviations of the names of corresponding content elements. Since it is very difficult to judge in advance which content elements would be the most valuable for a particular user and which would not, it seems reasonable to require the enumeration of all content elements occurring in the text of the document abstracted.
4. AN EXAMPLE OF THE PS ABSTRACT

Below is an example of the complete PS abstract prepared for the document [29]:

*Document problem:*

What is the model of an information retrieval system that is intermediate between Boolean and vector-processing one?

*Problem solution:*

The extended Boolean model is presented. It preserves the query structure inherent in the Boolean model while at the same time it allows the weighted terms to be incorporated both in queries and stored documents. The extended system produces better retrieval output than either the Boolean or the vector-processing system.

*Testing method:*

The extended system has been compared with the Boolean and the vector processing one in respect of recall and precision. 4 files of documents on biomedicine, library science electrical engineering and computer science containing correspondingly 1,033, 1,460, 12,684 and 3,204 documents have been searched.

*Content elements:*

ORI DEF SYM THY SOA LIM SOL MAT ILU CHS IDT SCT PLT CRI PRO JUP
OBJ CHO SOO PRE CAL TAB

5. TESTING THE METHOD OF WRITING PS ABSTRACTS

The testing method consisted in preparing PS abstracts of 80 documents from information science and comparing them with abstracts of the same documents included in the INSPEC database. The content and length of abstracts were used as the main evaluation criteria. The results of comparison are shown in Table 1.

In order to compare PS abstracts with INSPEC abstracts the answers to the following questions were found: “Were document problem, problem solution, testing method and related problems reflected in a given abstract?” In the case of the testing method the idea of the method, the scope of testing and the evaluation criteria were taken into account. Each of the above questions were given one of three possible answers: yes, partially, no.

As shown in Table 1, PS abstracts contained formulations of all problems which had been solved and then described in a document. 49 percent of problem solutions were presented in PS abstracts completely and 34 percent of problem solutions—only partially. The reason for this is as follows: In the case of the problems of the type “what?,” e.g., “what reason . . . ?,” “what language . . . ?,” “what properties . . . ?,” the answers presented in the primary documents were comprehensive and rather voluminous. An attempt to describe them to the full extent or to enumerate all elements could increase the length of the PS abstract to such an extent that it would exceed the reasonable size. The analogous situation occurred in the case of “how?” problems. However, besides the necessity of limiting
the abstract length another factor played the role. In a few instances the property character-
izing the method could not have been determined. In PS abstracts, (partially descri-
bing the solutions of "how?") problems of the principles on which the methods were based,
or specific aspects of the usefulness of the methods, were presented. Limiting the length
was also the reason for presenting the evaluation criteria and the scope of testing in the
PS abstracts partially.

On the basis of data shown in Table 1 it may be concluded that INSPEC abstracts give
less information about the problem solving process than the PS abstracts do.

The lengths of the PS abstracts written in Polish and the INSPEC abstracts translated
into Polish were determined in characters. The comparison of the lengths revealed that 53
PS abstracts were shorter and 26 were longer than the corresponding INSPEC abstracts.
In one instance, the lengths were equal. The total length of 80 PS abstracts amounted to
44,069 characters (without content elements parts) and of 80 INSPEC abstracts—to 51,001
characters. Thus the total length of PS abstracts was 85 percent of the total length of
INSPEC abstracts. Taking into account the fact that the total length of PS abstracts does
not cover the content elements parts, this length may be considered comparable to the total
length of the INSPEC abstracts.

6. CONCLUSIONS

In this paper a methodology for content analysis of scientific documents and for writ-
ing PS abstracts has been introduced. Taking the problem solving process as the main sub-
ject of a document and incorporating it into a structured abstract makes it possible to
direct information retrieval towards particular stages of scientific research. The structure
of the PS abstract and the content elements enable the scientist to concentrate on the in-
teresting aspect of the document content and to select documents satisfying his/her indi-
vidual information needs easily. This will be possible because PS abstracts reflect the
structure of needs and preferences of a scientist's solving problems.

The method presented in this paper facilitates identifying and selecting the most
important information from a document. The rules of how to present this information in
a PS abstract allow one to obtain abstracts which are highly informative and uniform.
Testing the method showed that PS abstracts as compared with INSPEC abstracts carry
more information both about the problem solving itself and about the manner and scope
of its presentation in a document. Moreover PS abstracts have comparable length with the
length of INSPEC abstracts. The disadvantage of the PS abstract is that its transparency
decreases when the number of described problems (sub-problems) increases.

PS abstracts could be used in scientific journals and other publications containing
original scientific papers or reports as well as in automated retrieval systems. In the former
case the rules of writing PS abstracts and explanations of content element codes should be
placed in each issue of a particular journal, e.g. on its cover. Using PS abstracts in an auto-
mated information retrieval system, particularly in online systems, might enable the user
to look through a great number of abstracts easier and faster. It could be possible to pro-
vide the user with only those parts of PS abstracts which were specified by him/her in a
query. Thus, some initial automatic reduction of the search output according to the users'
needs and preferences might be accomplished.

REFERENCES

1. Witty, F.J. The beginnings of indexing and abstracting: Some notes toward a history of indexing in antiq-
6. Buxton, A.B.; Meadows, A.J. Categorization of the information in experimental papers and their author
A methodology for writing problem structured abstracts


APPENDIX

The list of the content element codes to be used in the PS abstract follows. Frequency numbers of each content element in the set of 80 articles are given in parentheses.

REC — receiver of document (5)
SCO — scope of document (6)
ORI — origin of document problem (54)
DEF — definitions of concepts (42)
SYM — explanation of symbols (32)
THY — theories, principles used (41)
HIS — historical presentation of relevant research (2)
SOA — state of the art (46)
KNO — description of known solution (5)
LIM — limitations, assumptions of problem (31)
HYP — hypotheses proposed (2)
SOL — description of problem solution (57)
IDS — idea of solution (13)
ALG — presentation of solution in form of an algorithm (17)
MAT — mathematical description (28)
THM — theorems (16)
PRF — proofs of theorems (16)
ILU — illustration of solution (47)
JUS — justification of solution (19)
CHS — characteristics of solution (36)
IDT — idea of testing method (51)
SCT — schedule of testing method (15)
MOD — model used (1)
JUT — justification of testing method (3)
PLT — place where testing was carried out (17)
TIT — time of testing (12)
CRI — evaluation criteria used (37)
CHC — characteristics of criteria (4)
JUC — justification of criteria (5)
PRO — specification of procedures employed in testing (20)
CHP — characteristics of procedures (10)
JUP — justification of procedure selection (6)
SOP — source of procedures (0)
OBJ — specification of objects used in testing (31)
CHO — characteristics of objects (15)
JUO — justification of object selection (6)
SOO — source of objects (e.g., producer or supplier) (15)
EQU — specification of equipment used (15)
CHE — characteristics of equipment (3)
JUE — justification of equipment selection (1)
SOE — source of equipment (e.g., producer or supplier) (1)
PRE — preliminary activities (12)
SID — side effects (2)
RAW — presentation of raw data obtained (2)
CAL — data reduction, calculations (5)
RES — presentation of results of calculations (22)
TAB — presentation of results in form of tables (26)
FIG — presentation of results in form of figures (8)
SCH — presentation of results in form of schemes, maps (4)
CPL — evaluation of data completeness (3)
PRC — evaluation of data precision (2)
ERR — analysis of possible errors (1)
STA — statistical analysis of results (9)
NEG — description of negative results obtained (12)
EXP — explanation of results obtained (16)
GEN — generalizations (23)
COM — comparison with results obtained by other authors (31)
USP — possible usage areas in practice (39)
USS — possible usage areas in science (11)
IMP — possible ways of improving solution (30)
NEW — new problems encountered during research (2)
COS — costs of research (1)
FIN — financial sources (19)