Hans H. Wellisch
College of Library and Information Services,
University of Maryland

Some vital statistics in abstracting and indexing

Wellisch, H. H.: **Some vital statistics of abstracting and indexing.**
In: Intern. Classificat. 7 (1980) No. 3, p. 135–139

The almost 2400 items listed in a recently published bibliography of abstracting and indexing (A&I), covering a period from 1856 to 1976, are analyzed by number, age, language, and subject matter. About 90% of the A&I literature has been written since 1950. English is the predominant language (83%), Russian, German, French and Japanese contributed 15%, and all other languages together 2%. Russian material is on the rise, and will in future contribute about 10% of the A&I literature, mostly on automatic methods. The literature on automatic A&I is steadily growing and constitutes more than 15% of all entries. The rise and fall of some topics as measured by contributions throughout the period covered is traced. A&I methods invented in one country took about 8–10 years to be adopted in other countries, but the adoption rate is now somewhat faster, especially regarding automatic methods. (Author)

1. General Description of the Bibliography

The compilation of an extensive bibliography on abstracting and indexing (A&I) (1) affords an opportunity to examine some trends in the literature that deals with the focal concern of information science, namely the indication of the subject content of documents by various methods. Before attempting to analyze the data and to draw some conclusions based on the number, age, language, and to some extent also on the subject matter of entries in the bibliography, it is necessary to state its scope and limits. The entries are primarily on topics pertaining to the indication of subject content by means of *words, terms* or *phrases*, but exclude most of the writings on systematic subject indication, i.e. by classification. (The latter will be covered by a planned bibliography recently announced in this journal[1].) However, comparisons between verbal indexing systems and classification, as well as contributions on the indexing of classification schemes are included. The bibliography also includes items on alphabetical filing because all indexing depends for its effectiveness on a predictable order of entries. The subject of thesauri, which is well covered by several bibliographies[2], was included only inasmuch as items dealt with thesauri in relation to indexing, with terminological questions, and with the use of thesauri by indexers. Excluded were most items which dealt only with a specific local application of an indexing system without contributing anything substantially new to the state of the art, that is to say, the type of article best characterized by "How I indexed our collection of ..." Some contributions from countries outside the English-speaking world which would perhaps have been repetitious had they been written in English (the predominant language of the items listed) were included because they show when certain indexing methods, mostly developed in the U.S. or U.K., reached users in other countries, and how they were applied to tasks that are sometimes specific to the information needs and systems of those countries. These non-English entries are written in 24 languages, as shown in Table 1.

Table 1: Distribution of entries by language

Language	*Items*	*Language*	*Items*
Afrikaans	1	Hebrew	5
Arabic	1	Hungarian	9
Bulgarian	2	Italian	4
Chinese	4	Japanese (JPN)	30
Czech (CZE)	22	Norwegian	5
Danish	8	Persian	1
Dutch (DUT)	13	Polish (POL)	13
English (ENG)	1990	Portuguese	3
Finnish	1	Rumanian (RUM)	20
Flemish	1	Russian (RUS)	110
French (FRE)	40	Spanish	2
German (GER)	101	Swedish (SWE)	10
		Ukrainian	2

The total number of entries is 2398, whereas in Table 2 the total is only 2362, because the items published in 1977 and 1978 have not been included. The three-letter language codes (according to *MARC Language Codes*, 1980) refer to the language abbreviations used in Table 2.

2. Period Covered

Although A&I has a long history, dating back at least to the 4th century B.C. in Greece, and possibly even to earlier periods (the brief indications of the contents of cuneiform tablets, written on their edges or on clay envelopes, probably served an indexing function), the practitioners of the art seldom if ever wrote on its techniques until about a hundred years ago. One major and truly remarkable exception was the Swiss polyhistor and founder of modern bibliography, Conrad Gessner. He was the first to give explicit theoretical and practical advice on the making of indexes in one of the introductory chapters of *Pandectarum ... libri XXI*, the huge classified subject index to about 12,000 books listed in his earlier *Bibliotheca universalis* (Tiguri: Froschauer, 1548–49)[3]. But apart from some brief notes on alphabetical indexing and filing found here and there[4], nothing noteworthy on indexing appeared until the mid-19th century, when Andrea Crestadoro's famous pamphlet *The Art of Making Catalogues for Libraries* was published in 1856. Since this work anticipated what more than a hundred years later became known as KWIC indexing, and is therefore of more than curiosity value, it has been included as the oldest item in the bibliography. Although Crestadoro's ideas are said to have been realized in the catalog of the Manchester Public Library, they did not have any further impact on subject indexing (probably because the necessary technology was not yet available). Apart from a few articles in various

https://doi.org/10.5771/0943-7444-1980-3-135 https://www.inlibra.com/de/page/agb - Open Access -

literary magazines on the necessity of indexes or the woeful inadequacy of those that were produced at the time, professional writings on indexing did not really begin to appear until the foundation of the *Library Journal* by Melvil Dewey and Charles Ammi Cutter in 1876. Thus, although the bibliography technically covers a period of 120 years, it is more correct to say that it is a record of exactly 100 years of writings on A&I, the cutoff date for entries being 1976. (There are also a few entries for monographs published in 1977 and 1978, but these have not been taken into consideration in the following computations.)

3. Distribution by Period

Table 2 shows that almost 90% of the total literature on A&I has been written in the post-World War II era, and more than three quarters date from the period beginning in 1960. Thus, sustained and increasingly intensive professional interest and study of the various aspects of A&I is a fairly recent phenomenon, not older than about 25 years.

4. Distribution by Language

Table 2 also shows that English is the predominant language of writings on A&I, not only during the earlier period from 1876 and until the 1930's (that is, at a time when the sciences were dominated by contributions in German and French) but also consistently throughout the entire period covered. English items constitute 83% of the total (1954 out of 2362), although non-English contributions, as we shall presently see, are now steadily on the rise. In this respect, A&I seems to be characteristic for the whole field of library and information science which, despite a growing number of contributions from the non-English-speaking world, is a field dominated by English to an extent probably found in only few if any branches of learning.

Regarding non-English languages, the earliest contributions listed were written in the early or mid-1930's when a few German, French, Dutch and Norwegian articles began to appear, but a glance at Table 2 shows that, similar to the English contributions, those in other languages became substantial only after 1950. At this point, it must be conceded that the coverage of German and French material from earlier periods is not as thorough as that for English material, primarily because the indexing of periodical articles in these and other European languages before 1930 is spotty at best and often non-existing. It stands to reason that, to mention just one example, the *Zentralblatt für Bibliothekswesen*, founded in 1886, probably published some articles on indexing and alphabetic filing long before 1933 (when it was first indexed by *Library Literature*), but the relatively small number of English writings on A&I during the same period makes it appear unlikely that the inclusion of early articles in other languages would have made a major difference in the overall picture. In this context it is perhaps significant to note that the first English textbook on indexing (4) appeared in 1878, while the first German one (5) was published only in 1964.

Altogether, there are 408 non-English items, or 17% of the total. Most of these, especially in the less well-known languages, are introductions and summaries of

Table 2: Distribution of entries by period and language

Period		ENG	RUS	GER	FRE	JPN	CZE	RUM	DUT	POL	SWE	Other	Subtotal	Total	%	Cum.%
1856–1875	I	5											5	5	0.0	
	A	1														
1876–1899	I	38											38	39	1.7	1.7
	A	1											1			
1900–1924	I	48											48	56	2.4	4.1
	A	8											8			
1925–1949	I	92	2	1	8				5			5	113	147	6.2	10.3
	A	32			1				1				34			
1950–1954	I	66		10	3						1	3	83	99	4.2	14.5
	A	15							1				16			
1955–1959	I	109		18	2				2		3	2	136	169	7.2	21.7
	A	31		2									33			
1960–1964	I	283	3	17	3	1		1			4	5	317	380	16.1	37.8
	A	55	3	2	1			1	1				63			
1965–1969	I	372	20	16	6	12	1	7	1	3	1	10	449	557	23.6	61.4
	A	88	5	3	1	5		1	1	4			108			
1970–1974	I	459	41	20	11	9	16	9		5	1	11	582	668	28.3	89.7
	A	63	14	5	1	2			1				86			
1975–1976	I	163	16	7	4	1	4	1		1		12	209	242	10.3	100.0
	A	26	6				1						33			
Subtotal		1635	319	82 28	89 12	37 4	23 7	21 1	18 2	8 5	9 4	10	48	1980	382	
Total		1954		110	101	41	30	22	20	13	13	10	48	2362	2362	
Percent		82.7		4.7	4.3	1.7	1.2	0.9	0.9	0.6	0.6	0.4	2.0	84	16	

I = Indexing. A = Abstracting. Language codes: see Table 1.

https://doi.org/10.5771/0943-7444-1980-3-135 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

methods developed primarily in the English-speaking countries, and illustrate their adaptation to conditions in different lingual environments, and their applications to the exploitation of local collections. In some instances, notably in the case of Czech and Romanian, a large number of contributions are the work of one or two individuals who consistently write on indexing problems.

Table 2 also shows that writings on indexing far outweigh those on abstracting (although there are many articles that treat both topics in conjunction, and these are counted only under indexing). It seems that apart from English, Russian, and German authors writers in other languages are only marginally interested in the topic of abstracting.

Regarding a detailed breakdown of items by language, Table 2 specifies only the contributions of the ten languages that produced at least 10 items over the whole period. The first five happen to correspond exactly to the ranking of scientific articles by country of origin in the *Science Citation Index*[5]. The production of languages with less than ten items is summarized under "Other" and constitutes a mere 2% of the total. It is easy to discern that there are essentially only three top producers of A&I literature, namely English, Russian, and German. Table 3 analyzes the contributions of these languages, in column A for the entire period, in columns B and C for the past 17 and 7 years respectively. Whereas the contributions of the three languages produce fairly consistently about 90% of all material, the percentage of Russian material is steadily rising, but the German one, after a rise in the 1960's (due in large part to a considerable number of articles and books on edge-notched and optical coincidence cards) seems to be on the decline during the 1970's. It is probably safe to predict, based on the figures in Table 3 and on the number of items that have appeared in the meantime until the summer of 1980, that Russian items will in future constitute about 10% of all A&I material (most of it devoted to automatic methods), although these are often rather brief articles of no more than a few pages, and many are of local interest only or are dealing with problems in highly specialized technical fields of information.

Table 3: The three top producers of A&I literature

	A 1856–1976		B 1960–1976		C 1970–1976	
	N	%	N	%	N	%
English	1955	90.3	1169	86.8	711	86.7
Russian	109	5.0	108	8.0	77	9.4
German	101	4.7	70	5.2	32	3.9
Total	2165	100.0	1347	100.0	820	100.0
All entries	2362	91.7	1506	89.4	910	90.1

Another language whose production of A&I material is on the rise is Japanese whose share might well reach 2% of the total in years to come, equaling that of all other minor languages combined. Because of the difficulty of this language and the obvious concentration of Japanese authors on the specific problems caused by indexing in the Japanese writing system, these contributions will however be of only marginal interest for Western users. Lately, some Japanese investigations of a more general nature have been translated by their authors into English, and have been published in English-language journals, thus following an accepted trend in scientific and technical communication.

5. The Rise and Fall of Some Topics in A&I

Since all items in the bibliography are arranged chronologically under each topic and subtopic, it is easy to follow the development of ideas, and in certain instances the rise and fall of systems and methods. Unlike many other subjects of study which during their early period were being dealt with in a manner that to later generations seems to be simplistic or pedestrian, or at best of historical value only, some of the earliest writings on indexing could almost have been written today. Much of their practical advice on, say, the indexing of names is as valid today as it was a hundred years ago. Likewise, one may find that SDI services (though as yet not dignified by that acronym) were apparently flourishing in American special libraries around 1914, complete with abstracts which were circulated to management and staff of a technical firm, and then indexed in the library for future reference. But only very few of the early writers on indexing gave much thought to the theoretical foundations of their craft. For these, one must turn to the contributions of the last two decades or so, when the nature of indexing, its intellectual processes, and its relationship to linguistics and classification have been explored.

On the other hand, many authors who claimed to have discovered the philosophers' stone of indexing, and to have found the hidden truth on how indexers and abstracters go about their work, did apparently not bother to read what others before them had said, and went blissfully on to reinvent the wheel. This phenomenon is particularly prevalent in some of the entries in which new automatic indexing techniques are revealed to the world. Often quite unreasonable claims are being made, while cost figures, so important in the real world as opposed to what can be done experimentally, are seldom if ever given. Be that as it may, items on automatic methods of A&I show a dramatic increase from a few scattered references in the early 1950's to several dozen contributions in many languages during each of the last ten years covered. (For the following computation, only items dealing with the automatic generation of index entries and/or abstracts have been taken into consideration. Items on the physical production of indexes by computer-controlled methods, i.e. alphabetical arrangement and printing of entries, whether humanly produced or machine-generated, are not included.)

Table 4 shows that there are 367 entries, or 15.5% of the total, and the figures indicate that the number of contributions is steeply rising. In this field, the percentage of non-English contributions (80 items) is much higher than for the subject of A&I as a whole, namely 22.6% for automatic A&I versus 17% for A&I in general. Of these, 40 items, exactly half of all foreign-language writings on automatic A&I, or 10.9% of all items on the topic, are in Russian. Table 5 shows a breakdown of articles on automatic A&I by topic and method. More than a quarter of all items deal with KWIC or KWOC indexing, the earliest and crudest, but still the most widely used form of automatic indexing. The methods listed are of course not mutually exclusive, and several

https://doi.org/10.5771/0943-7444-1980-3-135 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Table 4: The development of some topics in A&I

		Pre-1950	1950-1954	1955-1959	1960-1964	1965-1969	1970-1974	1975-1976	Total
Automatic indexing	E		1	7	59	65	95	29	256
	N				2	26	42	10	80
Automatic abstracting	E			4	8	8	4	1	25
	N					2	2	2	6
Total autom. methods			1	11	69	101	143	42	267
Chain indexing	E	2	4	7	5	4	17	3	42
	N							1	1
Total		2	4	7	5	4	17	4	43
Uniterms	E		7	13	16	5	3	2	46
	N			1	2	2		2	7
Total			7	14	18	7	3	4	53
Links and roles	E				11	13	4	2	30
	N				3		1	3	7
Total					14	13	5	5	37

E = English N = Non-English

Table 5: Items on automatic A&I, by method

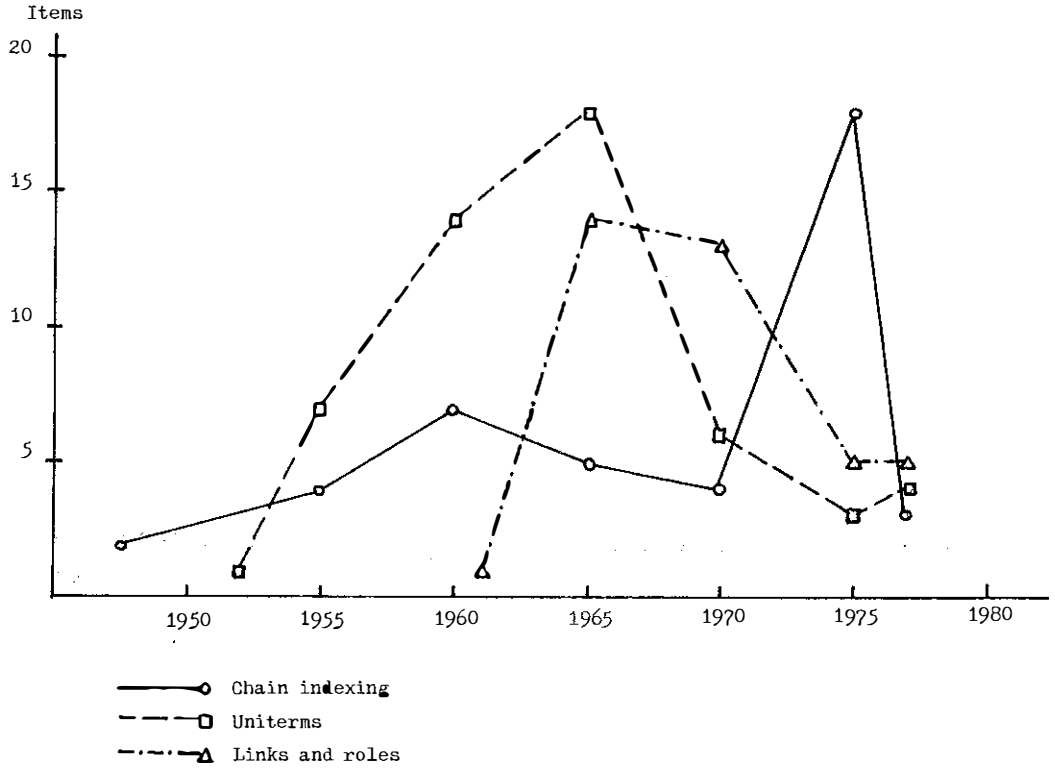
	N	%
General	56	15.3
KWIC & KWOC	93	25.3
Keyword extraction	67	18.3
Term frequency	57	15.5
Syntactic & semantic analysis	42	11.4
Associative methods	29	7.9
SMART	16	4.4
Term weighting	7	1.9
Total	367	100.0

of these are dealt with simultaneously in many of the contributions, but for the purpose of listing in the bibliography, each item appears only once, in the section that is most characteristic for it.

Reverting now to Table 4, I have chosen three other topics for closer analysis, namely Chain indexing, Uniterms, and Links and roles. These methods and techniques share the characteristic that they were invented by identifiable individuals, enjoyed a certain amount of popularity, resulting in many articles and reports, but declined in importance after 1975, as shown in Figure 1.

The case of Uniterms is perhaps typical for the dissemination and use of a method first invented in the United States, and then gradually adopted also in other countries. (For this computation, only articles explicitly dealing with Uniterms in their original or only slightly modified form were counted, but not the much larger number of contributions on coordinated indexing in general.) Of the writings on the topic in languages other than English, one is in Swedish in the period 1955–59, one each in German and Afrikaans in the period 1960–64, one each in Hungarian and Japanese in the period 1965–69, and

Figure 1: Number of entries published on three topics



https://doi.org/10.5771/0943-7444-1980-3-135 https://www.inlibra.com/de/page/agb - Open Access -

one each in Persian and Portuguese in the period 1975–76. Some of the English articles were also written by authors from non-English speaking countries, and these are occasionally published in journals which normally carry contributions in the respective vernacular. A case in point is the subject of Links and roles, whose heyday was between 1965 and 1969; from 1974 to 1976, only five items were written, and only two of these were in English (but written by a Danish and a Romanian author respectively), and one each was in Czech, French, and Russian. During that period, English-speaking authors had apparently lost interest in the subject.

Thus, a time lag of eight to ten years seems to be typical for methods and systems invented in one country or language area until their adoption and application in other countries, but this does not seem to hold true for automatic A&I methods which, as we have seen, are dealt with by authors from different countries almost as soon as they have been originated somewhere.

Sometimes, a certain national or regional bias for or against a certain method seems to be discernible, as in the case of manually operated optical coincidence cards (also known as "Peek-a-boo" cards) which found much use and generated a considerable literature in the United Kingdom and in Germany, but were almost entirely neglected in the United States, where only a mechanized version had a relatively brief period of limited use.

These are only a few examples, chosen at random, and I am sure that the bibliography will provide ample raw material for similar studies on other topics in the field of A&I.

The first supplement to the bibliography, which will cover the period 1977–1980 as well as additional items for the previous period, is now in preparation. The material already collected shows that the number of contributions to the field of A&I is still steadily growing, both in absolute numbers and in the part played by material originating in countries outside of the English-speaking world. It will be interesting to compare the demographic trends of the past with those of the four or five last years of the decade of the 1970's.

Notes:

1 See the Editorial of Intern. Classificat. 7 (1980) No. 1.
2 Two recent bibliographies, each containing about 800 entries, are (2) and (3).
3 My analysis and translation of this chapter, "How to make an index – 16th century style" will appear in a forthcoming issue of this journal
4 One of these, a compendium on "Useful and funny artful questions", will be discussed in an article "A German instruction in indexing from the 17th century" to be published in *The Indexer*.
5 The two top producers in the *SCI* are the U.S. and the U.K., whereas Canada and India follow Japan. In my bibliography there are however only a few Canadian and Indian items, and these have been subsumed under English or French.

References:

(1) Wellisch, H. H.: Indexing and abstracting: an international bibliography. Santa Barbara, CA: ABC-Clio Press 1980. 308 p.
(2) Aa, H.J.v.d.: Thesaurus bibliography. Amsterdam: Studiecentrum NOVI. 1977.
(3) MacCafferty, M.: Thesaurus and thesaurus construction. London: Aslib 1977.
(4) Wheatley, H. B.: What is an index? A few notes on indexes and indexers. London: Longmans, Green 1878.
(5) Kunze, H.: Über das Registermachen. Leipzig: Bibliographisches Institut 1964.

saur

The Capitals of Europe / Les Capitales de l'Europe

A Guide to the Sources for the History of their Architecture and Construction / Guide des Sources de l'Architecture et de l'Urbanisme

Hrsg. v. International Council on Archives
Redaktion: Agnes Ságvári

1980. 576 S. Ln. mit Schutzumschlag. DM 120,–
ISBN 3-598-10093-0
Englisch/Französisch

Ziel dieses Quellenbandes ist, grundlegende Informationen über das Wachstum und die Entwicklung der historischen Stadtkerne der heutigen Hauptstädte zu erteilen. Jede Hauptstadt wird in einem separaten Kapitel mit historischer Einleitung, Bibliographie, Illustrationen und Quellen behandelt.

Actes de la Seconde Conférence des Archives antillaises

(Guadeloupe et Martinique, 27-31 Octobre 1975)

Proceedings of the Second Caribbean Archives Conference

(Guadeloupe and Martinique, October 27-31, 1975)

Hrsg.: International Council on Archives with the financial aid of UNESCO

Archivum, Special Volume No. 2

1980. 160 S. Br. DM 48,–
ISBN 3-598-21002-7
Französisch/Englisch

Der vorliegende Band enthält Berichte und Arbeitspapiere der 2. Konferenz der Karibischen Archive vom 27.-31. Oktober 1975 in Guadeloupe und Martinique.

K·G·Saur München·New York·London·Paris

https://doi.org/10.5771/0943-7444-1980-3-135 https://www.inlibra.com/de/page/agb - Open Access -