3. INDEXES GENERATED BY MACHINE--AUTOMATIC DERIVATIVE INDEXING

We have noted, in the earlier statement of the scope of this survey, a distinction between "derivative" and "assignment" indexing. This distinction is related directly to the question: "Is what can be done by machine properly termed 'abstracting', 'indexing', or 'classifying'?" It relates also, as we have remarked, to a continuing controversy far older than any question of the introduction of machine techniques--that between "word" and "concept" indexing, between "uniterms" if selected directly from the text and "descriptors" in the sense of their being indexing terms selected so as to have "a carefully specified meaning for retrieval", [1] to say nothing of contrasts with subject heading schemes and classification schedules.

Some of the major arguments pro and con derivative (usually word) and assignment (usually concept) indexing will be considered in a subsequent section of this report on the problems of evaluating indexing methods. Nevertheless, the present popularity of automatic derivative indexes of the KWIC type, while subject to all the disadvantages typically cited for all purely derivative indexing systems, does show the actuality of automatic indexing potentialities and may in fact hold the promise of solving some of the present-day problems of subject control.

In this section, we shall consider first the straightforward word extraction techniques used in KWIC type indexes. Possibilities for modified derivative indexing by title augmentation, manipulation of word groups and use of special clues in keyword selection are then discussed, including work by Baxendale, Luhn, and Artandi. Related research and developments efforts work in automatic abstracting which lend themselves to derivation of indexing terms includes proposals and experiments by Luhn, Oswald, Edmundson, Wyllys, Doyle, and Lesk and Storm, among others. Some comments will be given on the quality of modified derivative indexing by machine. Automatic derivative indexing at the time of search, as in the natural language text searching systems of Swanson, Maron, Kuhns, and Ray, and Eldridge and Dennis, will be discussed in a later section of this report. [2]

3.1 KWIC Indexes

The development of computer-generated permuted-title keyword indexes, especially in the issuances of Chemical Titles and B.A.S.I.C. (Biological Abstracts-Subjects-In Context) has been hailed by some as "the miracle of the decade" and "the greatest thing to happen in chemistry since the invention of the test tube". [3] The major reason for the optimistic enthusiasm is the speed with which the computer can produce can produce a complete index to some specific set of books, documents or papers so that publication and dissemination of the index can be prompt and thus serve as an important tool in

---

1/ Mooers, 1963 [423], p. 3.

2/ See pp. 132-136.

3/ Quoted by D. R. Baker statement in "U. S. Congress, Senate Committee on Government Operations", 1960 [619], p. 169.

maintenance of truly current awareness. For example, Herner in his 1961 review of the state-of-the-art of organizing information says:

> "I am told that the American Chemical Society has never had a more successful basic science publication. The key to the whole thing is, I believe, the extreme currency of Chemical Titles. This in turn derives from the speed and simplicity of the KWIC process." 1/

Conrad reports as follows:

> "Reception of B.A.S.I.C. ... has been so extremely enthusiastic ... that we are excited by the possibilities of producing permuted title indexes in one or more additional languages. The creation of a B.A.S.I.C. index in any language requires only that the titles be translated and punched on cards. Alphabetical arrangement, permutation and 'type-setting' is completely automated and, for 5,000 titles takes only two hours to accomplish." 2/

3.1.1 Applications of KWIC Indexing Techniques

The KWIC type process is indeed simple and straightforward. The words of the author's title are prepared for input to the computer by keystroking, either to punched cards or to punched paper tape. After being read by the computer, the text of a title is normally processed against a "stop list" to eliminate from further processing the more common words, such as "the", "and", prepositions, and the like, and words so general as to be insignificant for indexing purposes, such as,"demonstration", "typical", "measurements", "steps", and the like. The remaining presumably "significant" or "key" words are then, in effect, taken one at a time to an indexing position or window, where they are sorted in alphabetical order. The result is a listing of each such word together with its surrounding context, out to the limit of the line or lines permitted in a given format. As each keyword is processed, the title itself is moved over so that the next keyword occupies the indexing position, and this process is repeated until the entire title has thus been cyclically permuted.

A number of formats are available in which the length of the line, the position of the indexing window, and the extent of "wrap-around" (bringing the end of a title in at the beginning of a line to fill space that would otherwise be left blank) are major variables. Current examples of KWIC type indexing output are shown in Figures 2 through 7. Usually, the indexing window is located at or near the center of the line with several extra spaces to the immediate left or with other devices such as the shading of B.A.S.I.C. to aid the searcher in scanning down the keywords listed. This is

---

1/ Herner, 1962, [266], p. 10.

2/ Conrad, 1962 [137], p. 378A.

VERVOLTAGE FOR HYDROGEN LIBERATION ON MANGANESE.= + ON THE O ZPKH-0035-2436
D CONTENT IN THE FODDER LICHENS OF THE TUNDRA.=+ASCORBIC ACI BOTZ-0047-1686
CHEMISTRY OF LICHENS. STRUCTURE OF UMBILICIN.= ACSA-0016-2240
NGSTEN-182.= HALF- LIFE OF THE 152 KEV TRANSITION IN TU PHYS-0028-1015
MEASUREMENT OF THE HALF LIFE OF THE 91 KEV STATE OF 147 PHYS-0028-1199
HAVING BIDENTATE LIGANDS. CHLORIDE ACETYL ACETONATES JCOM-0004-0481
OMPLEXES WITH DIFFERENT LIGANDS.= SPECTRA OF NICKEL(II) C MGKF-0068-0551
DISTRIBUTION OF LIGHT CHANGES OF CONTACT POTENTIAL FTVT-0004-3422
ON REGION IN IONIC + LIGHT DISPERSION IN EXCITON ABSORPTI FTVT-0004-3512
EFFECT OF OXYGEN ON THE LIGHT EMISSION PRODUCED BY DISSOLUTI ACSA-0016-2113
NCE OF POLARIZED HELIUM LIGHT EXCITED BY ELECTRONS.=+DEPENDE PHRV-0128-2222
LOGARITHMICAL OPTICAL LIGHT FILTER FOR QUANTITATIVE EMISSI MGFF-0010-0333
OF THE RESULTING LIGHT INDUCED BY PROTON IMPACT (5-35 PHYS-0028-1184
NCE + DECREASE OF THE LIGHT INTENSITY OF ELECTRO LUMINESCE APAH-0014-0155
ALS UNDER THE ACTION OF LIGHT IONS.= + EMISSION OF MET IANF-0026-1403
ELECTRIC+RECOMBINATION LIGHT OF INDIUM ANTIMONIDE IN STRONG APAH-0014-0245
LY PARA CHLORO STYRENE+ LIGHT SCATTERING AND VISCOSITY OF PO VMSD-0004-1839
LUTIONS.= LIGHT SCATTERING BY MIXED POLYMER SO VMSD-0004-1851
DETERMINATION OF THE LIGHT SCATTERING CONSTANT OF BENZENE DANK-0147-0573
IONS IN THE CASE OF LIGHT TARGETS BOMBARDED BY HEAVY IANF-0026-1440
SUBJECTED TO AGEING BY LIGHT.= + CHLORIDE) PLASTICS PLMS-62-11-059
EXCITATION BY MODULATED LIGHT.= + CRYSTALS IN THE CASE OF JINC-0024-0491
DI IODIDE) IN POLARIZED LIGHT.= + OF TRI(PHOSPHO NITRILIC ZENB-0017-0827
BY MEANS OF ULTRAVIOLET LIGHT.= + IN DEOXYRIBO NUCLEIC ACID DANK-0147-0227
APER CHROMAT+STRUCTURAL LIGNIN UNITS STUDIED BY METHODS OF P ACSA-0016-2303
ENYL)-2- PROPANONE FROM LIGNIN.=XY-1-(4- + PH IANK-1962-2094
RBONYL GROUPS OF COTTON LIGNINS.= + MOLECULAR WEIGHTS AND CA BCSJ-0035-2059
DETERMINATION OF LIGNO SULFONIC ACID IN SULFITE SPENT SSSA-0026-0574
FERTILIZATION. LIME AND P PLACEMENT EFFECTS ON JACH-0012-0535
REACTIONS IN HEATED LIME-ALUMINA MIXTURES.= JSSC-0013-0321
EFFECTS OF FERTILIZERS, LIME, AND CULTIVATIONS ON YIELD. ANYA-0102-0171A
ATION IN THE +PROBLEMS, LIMITATIONS, AND THE FUTURE OF AUTON JOSA-0052-1387
METRY.= DETECTION LIMITS IN RADIATION AND OPTICAL PYRO PLMS-62-11-053
WITHIN THE TEMPERATURE LIMITS 20-60-DEG.=+ GLASSY TEXTOLITE CORE-0255-3400
LEAR MAGNETIC RESONANCE LINE OF ZEOLITIC WATER.=+ OF THE NUC ACSA-0016-2212
MAGNETIC RESONANCE WIDE LINE SHAPES GENERATED BY TWO BROADEN ACSA-0016-2149
MAGNETIC RESONANCE WIDE LINE SPECTRA.= ANALYSIS OF NUCLEAR JCPS-0037-2699
N IODIDE.= LINE STRENGTHS AND WIDTHS IN HYDROGE FTVT-0004-3654
MAGNETIC RESONANCE LINE WIDTH IN GARNET AND SPINEL TYPE CANW-0007-0989
ORGANIC ELECTROLYTES IN LINEAR AND CIRCULAR CHROMATOGRAPHY JUPS-0017-1814
N-RANDOM DEGRADATION OF LINEAR CHAIN MOLECULES.= + OF A NO JUPS-0017-1694
N-RANDOM DEGRADATION OF LINEAR CHAIN MOLECULES.= + ON THE NO PHRV-0128-2131
THE ANTI FERRO MAGNETIC LINEAR CHAIN.= + SPECTRUM OF OR
DER-DISORDER THEORY FOR LINEAR COLLOIDS.= JCPS-0037-2323
OF THE COEFFICIENT OF LINEAR EXPANSION OF GLASSY PLASTICS. PLMS-62-11-036
OF THE SPECIFIC HEAT OF LINEAR POLYMERS AT LOW TEMPERATURES. DANK-0147-0580
INVESTIGATION ON THE LINEARITY OF THE TEST CURVE FOR THE ZACF-0192-0378
CHANGE OF K(35)- LINES AND THE VALUE OF THE INITIAL FMMT-0014-0660
RY OF SPECTRAL HYDROGEN LINES IN A PLASMA.= ASYMMET DANK-0147-0342
EUROPIUM(III) EMISSION LINES IN EUROPIUM DI BENZOYL METHANE JCPS-0037-2333
SITIES OF THE MANGANESE LINES.=+OF THE DIFFERENCE OF THE DEN MGKF-0068-0523
MMA-CONJUGATE SYSTEM OF LINKAGE.=+COMPOUNDS WITH A CLOSED GA DANK-0147-0636
CYANATE ON ETHYLENIC LINKAGES. ANALYTICAL APPLICATIONS. CHAL-0044-0463
INHIBITION OF ENERGY- LINKED DI PHOSPHO PYRIDINE NUCLEOTID JBCH-0238-0445
ODD NUMBER OF MONOMERIC LINKS, FORMED IN THE THERMAL POLYMER DANK-0147-0106
DDUCT+ISOLATION OF PURE LINOLENATE AS ITS MERCURIC ACETATE A JAOC-0039-0517
E+UREA FRACTIONATION OF LINSEED OIL FATTY ACIDS. COMPARATIV IJAC-0025-0053
CTIVITY OF LIPO PROTEIN LIPASE IN VARIOUS TISSUE SLICES.=+ A IGSB-0064-0105
IGGER ACTION OF PHOSPHO LIPASE ON A MAST, CELLS.= TR AIPT-0140-0577
PASE IN + LIPO PROTEIN LIPASE. ACTIVITY OF LIPO PROTEIN LI IGSB-0064-0105
LYCERIDES BY PANCREATIC LIPASE.= + ENZYMIC HYDROLYSIS OF G NMKK-0036-1004
T INTESTINAL + PHOSPHO LIPID COMPOSITION AND TURNOVER IN RA NATU-0197-0079
LIPID COMPOSITION OF TUMOR CELLS.= BIJO-0086-0350
OF GANGLIOSIDES IN LIPID EXTRACTS BY THIN-LAYER CHROMATOG BIJO-0086-0378
ON OF AN ORGAN SPECIFIC LIPID HAPTEN IN BRAIN.=+IDENTIFICATI NATU-0197-0060
STEROL AND PHOSPHO LIPID IN CEREBRO SPINAL FLUID AND CLCH-0008-0598
RE OF CERTAIN BACTERIAL LIPIDES.= + RELATED TO THE STRUCTU ANCP-0007-0593
AND METHIONINE ON SERUM LIPIDS AND LIPO PROTEINS.= PSEB-0111-0579
OF POLY SACCHARIDES LIPIDS AND NUCLEO PROTEIDS OF THE PEGT-08-06-075
METABOLISM OF PHOSPHO LIPIDS AS FUNCTION OF AGE AND UNDER ARBB-0005-0347
IDENTIFICATION OF LIPIDS IN BLOOD THROMBOPLASTIN.= PSEB-0111-0574
ERYTHROCYTE PHOSPHO LIPIDS IN THE NEWBORN INFANT.= PSEB-0111-0591
AND ADIPOSE TISSUE LIPIDS IN THE RATS RECEIVING HEPATO IGSB-0064-0036B
MYO-INOSITOL PHOSPHO LIPIDS OF MYCO BACTERIUM TUBERCULOSI JBCH-0238-0069
TION OF ERYTHRO+SPHINGO LIPIDS SERIES. SYNTHESIS AND RESOLU CCAC-0034-0161
INOVOSE FROM RED-CLOVER LIPIDS WITH ANTHRONE.=+ OF SULPHO QU NATU-0197-0062
HATE INTO MITOCHONDRIAL LIPIDS.= + INCORPORATION OF PHOSP JBCH-0238-0059
SEPARATION OF LIPO POLY SACCHARIDE AND MUCO PEPTID JBCH-0238-0026
OF LOW-DENSITY LIPO PROTEIN IMMUNO PRECIPITATES IN CLCH-0008-0616
LIPASE. ACTIVITY OF LIPO PROTEIN LIPASE IN VARIOUS TISSU IGSB-0064-0105
PO PROTEIN LIPASE IN + LIPO PROTEIN LIPASE. ACTIVITY OF LI IGSB-0064-0105
CONCENTRATION OF BETA- LIPO PROTEINS AND PROTEIN COMPOSITIO LABD-08-11-017
HIGHER FLOTATION CLASS LIPO PROTEINS AS THE CAUSE OF THE CCAT-0007-0872
SCLEHOSIS IN RABB+BETA- LIPO PROTEINS IN CHOLE STEROL ATHERO LABD-08-11-025
FOR DETERMINATION OF LIPO PROTEINS OF BLOOD SERUM BY LABD-08-11-021B
OD FOR DETERMINATION OF LIPO PROTEINS OF BLOOD SERUM.= METH LABD-08-11-021A
INE ON SERUM LIPIDS AND LIPO PROTEINS.= + AND METHION PSEB-0111-0579
ACTIVITY OF ADIPOSIN, LIPOCAINE AND PROLACTIN IN THEIR PEGT-08-06-037
EXERTED BY VITAMIN E ON LIPOIDS AND BLOOD COAGULABILITY IN VPIT-21-06-022
GOLD SENSITIZATION OF A LIPPMAN EMULSION.= PROBLEM OF THE ZNPF-0007-0465
NEW METHOD FOR STORING LIQUEFIED GASES.= ERKO-0015-0997
HYL ETHYL KETONE IN THE LIQUID AND GAS PHASES.= + OF MET KNKT-0020-0846
INFRARED SPECTRA OF LIQUID AND SOLID CARBON MON OXIDE.= JCPS-0037-2250
THE CASE OF STIRRING OF LIQUID AND SOLID PHASES.=+ACTION IN MGKL-0017-0466
ITY OF NEGATIVE IONS IN LIQUID ARGON, KRYPTON, XENON.=+MOBIL JCPS-0037-2470
EXPERIMENTS WITH GAS- LIQUID BUBBLERS.= AANL-0032-0358
C ACIDS-THEIR PREPARAT+ LIQUID C-18 SATURATED MONO CARBOXYLI JAOC-0039-0528
LECTRIC CONDUCTIVITY OF LIQUID CARBON-NICKEL ALLOYS.=+ AND E JANM-62-06-027
SYSTEM IN LIQUID- LIQUID CHROMATOGRAPHY.+ SEPARATION CRAB-0016-0901
ELECTROMOTIVE FORCE OF LIQUID COUPLES.= THERMO FMMT-0014-0789
US CARBON BLACK.+ LIQUID CRUDE FOR PRODUCTION OF GASEO GZVP-07-11-043
VISCOSITY OF LIQUID DEUTERO METHANE.= PHYS-0028-1197
DURING THE HEATING WITH LIQUID DI TOLYL METHANE.= KTTM-07-12-019
O EMISSION OF METALS TO LIQUID DIELECTRICS.= PHOT APAS-0015-0337
ECYCLE PROCESS. SOLID- LIQUID EQUILIBRIUM OF C8-AROMATICS.= SKGS-0005-0709
OF THE LAMBDA POINT OF LIQUID HELIUM IN THIN FILMS AND NUCI-0026-1010
SPECIFIC HEAT OF LIQUID HELIUM-3.= PHRV-0128-1983
PRESSURE OF HELIUM-3 IN LIQUID HELIUM-4, WITH PROPOSALS FOR PHRV-0128-1992
ATURE AND PRESS+FLOW OF LIQUID HELIUM(II) UNDER LARGE TEMPER APNY-0021-0072
DS IN THE RADIOLYSIS OF LIQUID HYDROCARBONS.= + RADICAL YIEL IJAR-0013-0493
ONDENSATION AT THE GAS- LIQUID INTERFACE.= POLYC VMSD-0004-1817
OF HYDROGEN IN LIQUID IRON BELOW THE BOILING POINT. DANK-0147-0628
N+DE PHOSPHORIZATION OF LIQUID IRON. EFFECT OF PHOSPHORUS O SRTA-0014-0316
E ACTIVITY OF OXYGEN IN LIQUID IRON.= + OF PHOSPHORUS ON TH SRTA-0014-0316
US. BY ALUMINOTHERMY OF LIQUID MANGANOUS SLAGS.= + PHOSPHOR ARRN-0007-0361
PURIFICATION OF LIQUID PARAFFINS WITH NITRO ETHANE.= KTTM-07-12-006
COBALTATE(II) ION. LIQUID PHASE HYDROGENATION AND DECOM JCTL-0001-0489
NS WITH METALLIC SALT + LIQUID PHASE OXIDATION OF HYDROCARBO TKSH-0057-0573
HENATE CATALYSTS ON THE LIQUID PHASE OXIDATION OF P-XYLENE.= TKSH-0057-0573
STATE OF THE VAPOR AND LIQUID PHASES IN PRODUCTS OF THE UFZH-0007-1136
ELECTRIC STUDY OF LIQUID SEMICONDUCTOR SOLUTIONS OF JCPS-0037-2677
WITHOUT CHANGING THEIR LIQUID STATE COMPOSITION.= BIOF-0007-0725
ITY OF TELLURIUM IN THE LIQUID STATE.= DENS CORE-0255-3406
NDIUM BISMUTHIDE IN THE LIQUID STATE.= + OF THE ALLOY I FMMT-0014-0787
XYL PHENYL KET OXIME IN LIQUID SULFUR DI OXIDE.=+OF CYCLO HE BCSJ-0035-1986
ALLY APPLIED DROPS OF A LIQUID SURFACE-ACTIVE METAL.=+OF LOC FMMT-0014-0757
TRANSFER BETWEEN GAS- LIQUID SYSTEMS AND THE HEAT EXCHANGE ZPKH-0035-2570
PROPERTIES OF THE LIQUID SYSTEMS ARGON-METHANE AND PHYS-0028-1191
IN A SYSTEM LIQUID- LIQUID WITH VARIOUS CONCENTRATIONS ZPKH-0035-2426
LOW NEUTRONS IN A FERMI LIQUID.= ON SCATTERING OF S NUPH-0040-0139

SOLVENT SYSTEM IN LIQUID-LIQUID CHROMATOGRAPHY.+ SEPAR CRAB-0016-0901
COEFFICIENT IN A SYSTEM LIQUID-LIQUID WITH VARIOUS CONCENTRA ZPKH-0035-2426
RELATIONS IN THE LIQUID-PHASE OXIDATION OF OLEFINS.= NEFT-0002-0750
SOLUTIONS IN DIPOLAR LIQUIDS BY THE METHOD OF MAGNETIC ZSTK-0003-0653
AND VISCOSITY OF LIQUIDS CONDENSED IN CAPILLARIES.= DANK-0147-0403
TING VAPOR PRESSURES OF LIQUIDS.= + METHODS FOR CALCULA MGKL-0017-0452
TION BANDS OF MOLECULAR LIQUIDS.= + OF VIBRATION ABSORP UFZH-0007-1090
C ACID IN SULFITE SPENT LIQUOR.= + OF LIGNO SULFONI BCSJ-0035-2059
SALTS IN SEWAGE-SLUDGE LIQUOR.=+ OF ORGANIC ACIDS AND THEIR ANAL-0087-0949
LE CRYSTALS OF ZINC AND LITHIUM ACETATES DI HYDRATES.=+ SING PISA-0056-0188
IES OF ALUMINUM-COPPER- LITHIUM ALLOYS.= + ON THE PROPERT IANH-62-06-150
PLANTS. ACTION OF LITHIUM ALUMINUM HYDRIDE ON THE YKKZ-0082-1652
CTRODE GLASS TO SILVER, LITHIUM AND THALLOUS IONS.= + ELE ANAL-0087-0977
M WITH BENZOIC ACID AND LITHIUM BENZOATE.= + ATOMS OF TRITIU RADK-0004-0615
RADIATION DAMAGE IN LITHIUM DOPED SILICON.= FTVT-0004-3373
RESSURE POLYMORPHISM IN LITHIUM HYDRIDE.= POSSIBLE LOW-P JCPS-0037-2730
EFFECTS OF SODIUM AND LITHIUM UPON THE RECEPTOR FOR THE SW AIPT-0140-0606
PROPANES WITH METHYL LITHIUM.= AN INTRA MOLECULAR ADDITIO CHIN-1962-2146
Y PHOTON BOMBARDMENT OF LITHIUM.=+BETA(-) ACTIVITY INDUCED B PRLT-0010-0055
AND CARBON-12 (LITHIUM-6,D) OXYGEN-16 FROM 3.4 TO PHRV-0128-2282
REACTIONS CARBON-12 (LITHIUM-6,P) OXYGEN-17 AND CARBON-12 PHRV-0128-2282
BROMIDE, POTASSIUM, LITHIUM, IODINE, RUBIDIUM IMPURITY FTVT-0004-3669
DETERMINATION OF LITHIUM, SODIUM, POTASSIUM, CALCIUM, CANW-0007-0943
ON CONDITIONS OF THE + LITHOLOGICAL COMPOSITION AND FORMATI DANK-0147-0170
TUBULAR REACTOR OF 18.1 LITRES CATALYST BED. REACTOR WITH TKSH-0057-0564
E ACTIVITY OF PINE-TREE LITTERS.= + AND FERMENTATIV PYDE-62-11-109
ACID COMPOSITION OF THE LIVER AND ADIPOSE TISSUE LIPIDS IN IGSB-0064-0034B
OF GUINEA PIG LIVER AND KIDNEY IN HISTAMINE SHOCK. BAPB-0010-0389
ODUCED BY CHLOROFORM IN LIVER AND KIDNEYS OF RAT.= + PR SPER-0112-0365
IN THE LIVER AND SPLEEN OF MICE AFTER INTRA BCPC-0011-1141
ENLARGEMENT OF THE LIVER AND SPLEEN OF THE O(20) (AMSTE BCPC-0011-1137
INYL PYRROLIDONE IN THE LIVER AND SPLEEN.= + AND POLY V BCPC-0011-1137
YDROXY METHYL + AZO DYE LIVER CARCINOGENESIS.= REACTION OF H NATU-0197-0087
LS IN THE PATIENTS WITH LIVER DISEASES.= + ACID LEVE JVIT-0008-0161
SIS OF ACETO ACETATE BY LIVER ENZYMES.= SYNTHE JBCH-0238-0064
ACID SYNTHESIS IN RAT- LIVER EXTRACTS. EFFECT OF X-IRRADIA BIJO-0086-0232
LEVELS OF DDT ON LIVER FUNCTION, CELL MORPHOLOGY, AND AIPT-0141-0111
WEIGHT OF CHICKEN LIVER GLUTAMIC ACID DEHYDROGENASE IN JBCH-0238-481P
AUTOMATION OF THE LIVER GLYCOGEN DETERMINATION.= ANYA-0102-0118
STERONE BY FEMALE-RAT- LIVER HOMOGENATE.= BIJO-0086-0365
ON OXIDATION BY LIVER HOMOGENATES ON MITOCHONDRIA OF BCPC-0011-1115
PHOSPHORYLATION IN LIVER MITOCHONDRIA FROM HYPOPHYSECTO JBCH-0238-0474
E ADMINISTRATION IN THE LIVER AND SPLEEN.=+ BY ETHONIN JBCH-0238-0469
ALTERATIONS IN THE LIVER OF WHITE RATS UNDER THE INFLUE FRMB-0010-0531
ATE ACYL TRANSFERASE IN LIVER PARTICLES.=+OF GLYCEROL PHOSPH BIJO-0086-0244
FREE FATTY ACIDS AND LIVER PATHOLOGY.. EFFECTS OF PHENOXY PSEB-0111-0731
PRODUCTION BY HUMAN LIVER SLICES IN VITRO BY AN IMMUNOPR JCIN-0042-0130
AND BUTYRATE BY SHEEP- LIVER SLICES.= + ACETATE, PROPIONATE BIJO-0086-0319
ACTIVITY IN KIDNEY AND LIVER TISSUE FROM STARVED AND FED PSEB-0111-0556
EXATE BY TRANSPLANTABLE LIVER TUMORS.= + CHLORO METHOTR PSEB-0111-0566
METABOLISM IN DISEASED LIVER.= SEROTONIN KLWO-0040-1233A
THE DEVELOPMENT OF RAT LIVER.= + ACTIVITY DURING NATW-0049-0609
VITY IN MOUSE BRAIN AND LIVER.= + ON GLUTAMINASE ACTI NATU-0196-1223
NTESTINAL TRACT AND THE LIVER.= + AMIDE IN THE GASTRO I YKKZ-0082-1588
NILIC ACID IN MAMMALIAN LIVER.= + FROM 3-HYDROXY ANTHRA JBCH-0238-483P
AL AND PRECANCEROUS RAT LIVER.= + OF AMINO ACIDS IN NORM PSEB-0111-0811
IN HEALTHY AND DISEASED LIVER.= + 5-HYDROXY- TRYPTOPHAN KLWO-0040-1229
IBO NUCLEIC ACID OF RAT LIVER.= + NUCLEOTIDES OF SOLUBLE R JBCH-0238-0311
FOR THE ACTIVATION OF LIVER-PYRUVATE OXIDASE SYSTEM RELATI JVIT-0008-0172
MIC NUMBER GREATE+HALF- LIVES OF FIRST EXCITED 2+ STATES ATO NUPH-0040-0024
Y BLOOD GAS EXCHANGE IN LIVING ANIMALS.= + OF INTRA PULMONAR JCIN-0042-0016
ICO TROPAL HORMONE, AND+ LOADING EXPERIMENTS WITH ADRENO CORT ENDK-0043-0155
UNDER HEATING AND UNDER LOADING.= + OF COPPER-NICKEL SYSTEM FTVT-0004-3409
OD BY PODOZOLIC MEDIUM- LOAMY SOILS.= + PROPERTIES OF S PYDE-62-11-017
HI AZOLIDINDIONES.+ NEW LOCAL ANAESTHETICS.= 3-ARYL-2,4-T BCSJ-0035-1926
OF DEUTERIUM OXIDE ON LOCAL ANESTHETIC ACTIVITY OF PROCAIN JPMS-0051-1166
NCREASING PRODUCTION OF LOCAL FERTILIZERS.= RESOURCES FOR I ZMLD-62-11-055
E BIOSYNTHESIS IN THE + LOCALIZATION AND SOURCES OF TREHALOS DANK-0147-0224
D INVERTASE ON SOLUBLE+ LOCALIZATION OF ACTION OF AMYLASE AN ARBB-0005-0399
IN THE PRESENCE OF LOCALLY APPLIED DROPS OF A LIQUID FMMT-0014-0757
TIN-BINDING SITES.= LOCATION AND ROLE OF STEROL AT NYSTA JOBA-0084-1152
LEIOTROPIC CARBOHYDRATE LOCUS IN STAPHYLOCOCCUS AUREUS.=+A P JOBA-0084-1338
INSECTICIDAL CONTROL OF LOCUSTS BY AERIAL SPRAYING.= CUSC-0031-0518
OR QUANTITATIVE EMISSI+ LOGARITHMICAL OPTICAL LIGHT FILTER F MGFF-0010-0333
S IN ITALIAN RYEGRASS (LOLIUM MULTIFLORUM).= + RELATIONSHIP NATU-0196-1230
DIRECT OBSERVATION OF LOMER DISLOCATION IN GERMANIUM CRYST JUPS-0017-1808
POLARIZABILITIES AND + LONGITUDINAL AND TRANSVERSE ELECTRON JCPS-0037-2421
HYSTERESIS LOOP OF COBALT-NICKEL FERRITES.= VMUK-0017-0058
ATION INTERACTION CONS+ LORENTZ PARAMETERS AND VIBRATION-ROT JOSA-0052-1374
AND CHROMIUM IN LOW ALLOY STEELS AND STAINLESS STEEL SRTA-0014-0294
ODINE-131 IN BITCHES ON LOW AND MODERATE FAT INTAKE.= + I PSEB-0111-0674
T YIELDS.= EFFECTS OF LOW DOSES OF GAMMA RADIATION ON PLAN IJAR-0013-0487
ESIUM-24 (P,P/GAMMA) AT LOW ENERGIES.= + MECHANISM IN MAGN PHRV-0128-2261
INELASTIC SCATTERING OF LOW ENERGY PROTONS FROM NEON-20.= NUPH-0039-0686
RUS-31 (P,GAMMA) SULFU+ LOW ENERGY RESONANCES IN THE PHOSPHO NUPH-0039-0631
GLANDS.+ EXTRACTION OF LOW MOLECULAR WEIGHT CORTICO TROPHIN ACEN-0042-0069
ID FROM C.+ISOLATION OF LOW MOLECULAR WEIGHT RIBO NUCLEIC AC SCIN-0028-0485
RIZATION OF ETHYLENE AT LOW PRESSURES.= METHOD FOR POLYME ANCE-0074-0977
CORTEX AS A CAUSE OF LOW RESISTANCE OF THE NEWBORN AND PEGT-08-06-008
N THE + INFLUENCE OF A LOW SODIUM DIET ON DYNAMIC CHANGES I VPIT-21-06-017
HELINES WITH CAUSTIC + LOW TEMPERATURE AGGLOMERATION OF NEP IVUT-05-05-093
AND + OXIDES FROM THE LOW TEMPERATURE OXIDATION OF NIOBIUM JCOM-0004-0533
EFORMED ALUMI+NATURE OF LOW TEMPERATURE TRANSFORMATIONS IN D FMMT-0014-0750
OF LINEAR POLYMERS AT LOW TEMPERATURES. VIBRATION SPECTRU DANK-0147-0580
ETIC BALANCE FOR USE AT LOW TEMPERATURES.= GOUY MAGN BCSJ-0035-2040
ON ADMIXTURE NUCLEI AT LOW TEMPERATURES.= MOSSBAUER EFFECT DANK-0147-0584
NUATION IN MAGNESIUM AT LOW TEMPERATURES.= ULTRA SONIC ATTE JUPS-0017-1801
DIOLYSIS OF N-HEXANE AT LOW TEMPERATURES.= + REACTIONS IN RA JCPS-0037-2496
NITRIDE INCLUSION IN LOW-CARBON-CHROMIUM STEEL.= IANH-62-06-073
MANGANESE REFINED FROM LOW-CONCENTRATION ORES CONTAINING ARRN-0007-0361
IPITATES +ESTIMATION OF LOW-DENSITY LIPO PROTEIN IMMUNO PREC CLCH-0008-0616
OF NEGATIVE HYDROGEN + LOW-ENERGY COLLISION CROSS SECTIONS JCPS-0037-2571
LEVELS OF MOLECULES BY LOW-ENERGY ELECTRON IMPACT SPECTROSC JCPS-0037-2497
ENERGY DISTRIBUTION IN LOW-ENERGY ELECTRON-PHOTON SHOWERS PHRV-0128-2352
ECTRUM OF INDIUM.= LOW-FREQUENCY LATTICE VIBRATIONAL SP JCPS-0037-2727
MIONIC CONVERTER WITH + LOW-FREQUENCY OSCILLATIONS IN A THER CRAB-0016-0889
GANIC COMPOUNDS USING A LOW-LEVEL NEUTRON SOURCE.= + IN OR BCSJ-0035-2045
DEUTEROLYSIS RATES OF LOW-MOLECULAR-WEIGHT ALKANE SULFONYL ZPCL-0221-0105
HYDRIDE.= POSSIBLE LOW-PRESSURE POLYMORPHISM IN LITHIUM JCPS-0037-2730
OF TYULENOVO OIL AFTER LOW-TEMPERATURE CATALYTIC TREATMENT. CRAB-0016-0949
WITH DI CHLORIDES FOR LOW-TEMPERATURE DEWAXING OF OILS.= KTTM-07-12-011
HYLENE.= LOW-TEMPERATURE POLYMERIZATION OF ET ANCE-0074-0955
ID HELIUM IN THIN FILM+ LOWERING OF THE LAMBDA POINT OF LIQU NUCI-0026-1010
6,7-DI METHYL-8-RIBITYL LUMAZINE UNDER ANAEROBIC CONDITION.= JVIT-0008-0178
ON COPPER(I) OXIDE LUMINESCENCE AND ELECTROCONDUCTIVITY FTVT-0004-3618
N THE CASE OF + PHOTO LUMINESCENCE AND VARIOUS PHENOMENA I APAH-0014-0247
HT INTENSITY OF ELECTRO LUMINESCENCE CELLS.= + OF THE LIG APAH-0014-0155
ODIUM IODIDE- + PHOTO LUMINESCENCE EXCITATION SPECTRA OF S UFZH-0007-1127
OXY LUMINESCENCE FROM COTTON TEXTILES.= TRJO-0032-1038
OCCURRENCE OF ELECTRO LUMINESCENCE IN ZINC SULFIDE SINGLE APAH-0014-0113
F + NARROW BANDS OF THE LUMINESCENCE OF COPPER BROMIDE AND O APAH-0014-0283
SODIUM IODIDE- + X-RAY LUMINESCENCE OF CRYSTALLO PHOSPHORS UFZH-0007-1101
OF OXYGEN ON THE THERMO LUMINESCENCE OF IRRADIATED POLY PRLA-0271-0188
LENE AND OTHER + THERMO LUMINESCENCE OF IRRADIATED POLY ETHY PRLA-0271-0170
E OF DEFECTS IN EXCITON LUMINESCENCE OF MOLECULAR CRYSTALS.= FTVT-0004-3634
POLARIZATION OF THE LUMINESCENCE OF PHENANTHRENE.= JCPS-0037-2413
HORS +EXTINGUISHING THE LUMINESCENCE OF ZINC SULFIDE LUMINOP APAH-0014-0105
CRYSTALS.+PHOTO ELECTRO LUMINESCENCE OF ZINC SULFIDE SINGLE APAH-0014-0121
CRYSTALS.= ELECTRO LUMINESCENCE OF ZINC SULFIDE SINGLE APAH-0014-0115
PYRROLE.= CHEMI LUMINESCENCE OF 2,3,4,5-TETRA PHENYL BCSJ-0035-2058

Figure 2. Sample Page, Standard IBM Format

Figure 3. Sample Page, B.A.S.I.C.

```
      FREQUENCY DOUBLING IN ANISOTROPIC FERRITES. ( SINGLE CRYSTAL ZINC(2)- YTTRIUM)                                  19-066
                        MAGNETIC SPIN PLANES IN MAGNETITE CRYSTAL.                                                    04-036
  MULTIPLE TWIN DOMAINS AND DOMAIN WALLS IN NICKEL- OXIDE CRYSTAL.                                                    06-062
PARAMAGNETIC RESONANCE OF THE COBALT ION IN RUTILE SINGLE CRYSTAL.                                                    12-046
AGNETIC ANISOTROPY MEASUREMENTS OF ANNEALED NICKEL- OXIDE CRYSTAL.                                                M   06-063
TUS FOR MEASURING MAGNETIZATIONS. APPLICATION TO A COBALT CRYSTAL.                                          A NEW APPARA  17-032
ESONANCE ABSORPTION OF DIVALENT NICKEL IN CORUNDUM SINGLE CRYSTAL.                                          PARAMAGNETIC R  12-016
LL ON SLOW NEUTRON SCATTERING BY A UNIAXIAL FERROMAGNETIC CRYSTAL.                                        EFFECT OF DOMAIN WA  01-070
FFECT AND THE ORDERING PROCESS IN A NICKEL(3) IRON SINGLE CRYSTAL.                                       MAGNETIC ANNEALING E  03-031
      MAGNETIC BEHAVIOR OF A TETRAGONAL ANTIFERROMAGNETIC CRYSTAL. ( THEORETICAL )                                    06-027
ISTRIBUTION OF DISLOCATIONS OVER THE CROSS SECTION OF THE CRYSTAL. /PART-2. EDGE AND SCREW DISLOCATIONS, D            04-073
RELAXATION OF TRIVALENT ERBIUM IN CADMIUM- IRON(2) SINGLE CRYSTAL. /RAMAGNETIC RESONANCE AND SPIN-LATTICE            12-057
EARTH-DOPED YTTRIUM IRON GARNET. / CONTRIBUTION OF STATIC CRYSTAL-FIELD EFFECTS TO THE LINE-WIDTH IN RARE-          11-020
OLYCRYSTALLINE MANGANESE- ZINC- FERROUS FE/ PERMEABILITY, CRYSTALLINE ANISOTROPY AND MAGNETOSTRICTION OF P         04-068
RITE- MAGNETITE AND MAGNESIUM FERRITE- MAGNETIT/ MAGNETIC CRYSTALLINE ANISOTROPY IN THE SYSTEMS NICKEL FER        04-147
ALS. ( LITHIUM(0.5)- ALUMINUM(2.5) OXYGEN(4) )  CRYSTALLINE ELECTRIC FIELDS IN SPINEL-TYPE CRYST                   04-091
D.                                  HYDROTHERMAL CRYSTALLIZATION OF YTTRIUM- IRON GARNET ON A SEE               18-003
 SOLUTION VANADIUM- OXYGEN(4)- COBALT(2-2X)- NICKEL (2X)/ CRYSTALLOGRAPHIC AND MAGNETIC STUDY OF THE SOLID       01-064
C PROPERTIES OF POTASSIUM MANGANESE(II) FLUORIDE. PART-1. CRYSTALLOGRAPHIC STUDIES.                    MAGNETI  05-035
ICROWAVE ACOUSTIC LOSSES IN YTTRIUM IRON GARNET. ( SINGLE CRYSTALS )          TEMPERATURE DEPENDENCE OF M    11-113
R- CHLORIDE DIHYDRATE, COBALT-CHLORIDE HEXAHYDRATE SINGLE CRYSTALS ) /IVITY IN AN ANTIFERROMAGNET. ( COPPE       06-050
/IENTATION AND ON THE METHOD OF DEMAGNETIZATION IN SINGLE CRYSTALS AND A POLYCRYSTAL OF 0.5PERCENT ALUMIN/       03-065
BALANCE FOR MEASURING ABSOLUTE SUSCEPTIBILITIES OF SINGLE CRYSTALS AND DILUTE SOLUTIONS. /SITIVE MAGNETIC        17-019
ON, AND PLASTIC DEFORMATION. COERCIVITY OF NICKEL SINGLE CRYSTALS AS A FUNCTION OF TEMPERATURE, ORIENTATI       03-007
          SYMMETRY OF TRANSITION METAL IMPURITY SITES IN CRYSTALS AS INFERRED FROM OPTICAL SPECTRA.                  16-031
             SPECIFIC HEATS OF SINGLE COPPER- MANGANESE CRYSTALS BETWEEN 1.4 AND 5K.                                16-029
                        GROWTH OF ALPHA- IRON SINGLE CRYSTALS BY HALOGEN REDUCTION.                                18-019
PART-1 A NEW METHOD OF PREPARING MAGNETITE SIN/ GROWTH OF CRYSTALS BY THE CHEMICAL TRANSPORT OF MATERIAL.         18-022
L/ MAGNETIZATION PROCESS IN UNIAXIAL FERROMAGNETIC SINGLE CRYSTALS FOR THE CASE OF A VERTICAL MAGNETIC FIE      02-097
ESE OXIDE, ALUMINUM OXIDE, MANGANESE SPINEL AND MAGNETITE CRYSTALS FROM 3 TO 300K. /CONDUCTIVITY OF MANGAN      16-027
TIONS.          GROWTH SEQUENCE OF GADOLINIUM-IRON GARNET CRYSTALS IN MOLTEN LEAD OXIDE- BORON- OXIDE SOLU       18-002
                    FORMATION OF MAGNETOPLUMBITE SINGLE CRYSTALS IN THE PRESENCE OF THALLIUM OXIDE.               18-021
RESONANCE TRIVALENT IRON AND DIVALENT MANGANESE IN SINGLE CRYSTALS OF CALCIUM OXIDE.        ELECTRON SPIN        12-030
.                 MICROWAVE RESONANCE LINEWIDTH IN SINGLE CRYSTALS OF COBALT-SUBSTITUTED MANGANESE FERRITE      11-081
IMENSIONS.    DEPENDENCE OF THE RESONANCE FIELD IN SINGLE CRYSTALS OF FERRITES ON TEMPERATURE AND SAMPLE D      11-032
/OF TITANIUM ON THE LOW TEMPERATURE TRANSITION IN NATURAL CRYSTALS OF HAEMATITE. ( ELECTRON SHADOW METHOD/       01-009
RIABLE WAVELENGTH.          MAGNETIC ANALYSIS OF SINGLE CRYSTALS OF IRON BY ELECTRON DIFFRACTION WITH VA       03-062
IATION WITH DEMA/ INITIAL PERMEABILITY OF SINGLE AND POLY CRYSTALS OF IRON- 5 PERCENT ALUMINUM AND ITS VAR      03-071
                    MAGNETORESISTANCE OF SINGLE CRYSTALS OF TRANSITION METALS.                                    09-006
                                        FERRITE CRYSTALS USING AN ARC IMAGE FURNACE.                              18-013
OPERTIES.                      THERMODYNAMIC THEORY OF CRYSTALS WITH FERROELECTRIC AND FERROMAGNETIC PR        02-095
                    DISLOCATIONS IN FERRITE SINGLE CRYSTALS WITH HEXAGONAL STRUCTURE.                             04-082
                ACOUSTIC PARAMAGNETIC RESONANCE IN CRYSTALS WITH IONS IN S-STATE.                                 12-002
              PHONON-MAGNON INTERACTION IN MAGNETIC CRYSTALS.                                                     01-021
       SYMMETRY PROPERTIES OF WAVE FUNCTIONS IN MAGNETIC CRYSTALS.                                                01-022
                DISORDER STRUCTURE IN TERNARY IONIC CRYSTALS.                                                     01-063
 X-RAY AND MAGNETIC STUDIES OF CHROMIUM- OXYGEN(2) SINGLE CRYSTALS.                                               01-065
   THEORY OF THE MAGNETIC SCATTERING OF SLOW NEUTRONS IN CRYSTALS.                                                 01-097
                MAGNETIC SPIN LEVELS IN MAGNETITE CRYSTALS.                                                       04-035
        NUCLEAR ORIENTATION IN ANTIFERROMAGNETIC SINGLE CRYSTALS.                                                 06-014
 THEORY OF NUCLEAR ACOUSTIC RESONANCE LINE SHAPE IN CUBIC CRYSTALS.                                               11-115
              ON MAGNETIC RESONANCE SATURATION IN CRYSTALS.                                                       12-008
 PARAMAGNETIC RESONANCE OF NICKEL IONS IN DOUBLE- NITRATE CRYSTALS.                                               12-036
       ASYMMETRIC SHAPE EFFECTS IN DIA- AND PARAMAGNETIC CRYSTALS.                                                14-015
            GROWTH OF YTTRIUM-ALUMINUM GARNET SINGLE CRYSTALS.                                                    18-001
  RESEARCH AND DEVELOPMENT OF YTTRIUM IRON GARNET SINGLE CRYSTALS.                                                18-015
                GROWTH OF REFRACTORY OXIDE SINGLE CRYSTALS.                                                       18-020
               GROWING YTTRIUM IRON GARNET SINGLE CRYSTALS.                                                       18-024
IFFUSION OF IRON AND CHROMIUM IN CORUNDUM AND RUBY SINGLE CRYSTALS.                                         D     12-032
 EFFECT OF SIXTH DEGREE CUBIC FIELD ON RARE-EARTH IONS IN CRYSTALS.                                        THE     14-040
ALENT CHROMIUM AND IRON RELAXATION TIMES IN RUTILE SINGLE CRYSTALS.                                       TRIV    12-031
WAVES IN RHOMBIC ANTIFERROMAGNETIC AND WEAK FERROMAGNETIC CRYSTALS.                                       SPIN    06-005
C INTERACTION OF CERIUM AND COBALT IONS IN DOUBLE NITRATE CRYSTALS.                                     MAGNETI   05-038
IC DOMAIN PATTERNS ON NICKEL-COBALT ALLOY AND PURE COBALT CRYSTALS.                                  FERROMAGNET  10-015
NEALING EFFECT ON THE ANISOTROPY OF COBALT FERRITE SINGLE CRYSTALS.                                  MAGNETIC AN  04-108
RESONANCE OF TRIVALENT IRON IONS IN SYNTHETIC ZINC- OXIDE CRYSTALS.                                 PARAMAGNETIC  12-024
ANCE OF DIVALENT MANGANESE IONS IN SILVER CHLORIDE SINGLE CRYSTALS.                          PARAMAGNETIC RESON  12-044
ATTERNS ON TWO-PHASE NICKEL- COBALT ALLOY AND PURE COBALT CRYSTALS.                        FERROMAGNETIC DOMAIN P  10-022
 OF TRIVALENT IRON IONS IN SYNTHETIC CUBIC ZINC- SULPHIDE CRYSTALS.                         PARAMAGNETIC RESONANCE  12-025
CTRON NUCLEAR DOUBLE RESONANCE OF PARAMAGNETIC DEFECTS IN CRYSTALS.                      FREQUENCY SPECTRA OF ELE  12-014
PY OF THE FERROMAGNETIC PRECIPITATE IN GOLD-NICKEL SINGLE CRYSTALS.          OBSERVATION BY ELECTRON MICROSCO  05-022
UND-STATE POPULATION CHANGES OF NEODYMIUM IN ETHYLSULFATE CRYSTALS.      DIRECT OPTICAL DETECTION OF THE GRO  14-012
   CREEP AND BASCULATION EFFECTS IN IRON- ALUMINUM SINGLE CRYSTALS. ( DEFECTS )                                   03-073
) )            CRYSTALLINE ELECTRIC FIELDS IN SPINEL-TYPE CRYSTALS. ( LITHIUM(0.5)- ALUMINUM(2.5) OXYGEN(4    04-091
                  ELASTORESISTANCE EFFECT IN IRON SINGLE CRYSTALS. ( MAGNETOSTRICTION ).                          03-043
STARK EFFECTS AND SPIN-PHONON INTERACTION IN PARAMAGNETIC CRYSTALS. ( THEORETICAL )                               13-005
LORIDE FROM 11 TO 300K. MAGNETIC ORDERING IN LINEAR CHAIN CRYSTALS. /AND ENTROPY OF COPPER AND CHROMIUM CH      16-023
SORPTION AND MANGANESE- MAGNESIUM- COBALT- FERRITE SINGLE CRYSTALS. /L POWER FOR THE CASE OF SUBSIDIARY AB      11-082
 THE FERRIMAGNETIC RESONANCE LINEWIDTH OF LITHIUM FERRITE CRYSTALS. /L, THERMAL, AND CHEMICAL TREATMENT OF      11-089
TERIAL. PART-1 A NEW METHOD OF PREPARING MAGNETITE SINGLE CRYSTALS. /STALS BY THE CHEMICAL TRANSPORT OF MA      18-022
 ON THE MAGNETIC DOMAIN STRUCTURE OF IRON- SILICON SINGLE CRYSTALS. /TERNAL STRESSES AND OF FIELD STRENGTH      10-017
PORATION OF ALPHA- HEMATITE INTO MANGANESE FERRITE SINGLE CRYSTALS. EFFECT ON DISLOCATION DENSITY.   INCOR      04-025
/FECTS IN YTTRIUM- IRON AND GADOLINIUM-IRON GARNET SINGLE CRYSTALS. PART-1. ETCHING AGENTS FOR GARNETS, O/      04-012
 LOW-INDEX FACE/ DISLOCATIONS IN MANGANESE FERRITE SINGLE CRYSTALS. PART-1. OBSERVATION OF DISLOCATIONS ON      04-072
ISTRIBUTION OF / DISLOCATIONS IN MANGANESE FERRITE SINGLE CRYSTALS. PART-2. EDGE AND SCREW DISLOCATIONS, D      04-073
TRIC PROPERTIES.                    SYMMETRY OF CRYSTALS, EXHIBITING FERROMAGNETIC AND FERROELEC  01-024
LD SPLITTINGS OF DIFFERENT IRON COMPLEXES. ( PARAMAGNETIC CRYSTALS, GARNETS )                  ZERO FIE  12-015
 OF ORIENTED NUCLEI. ( FERROMAGNETIC OR ANTIFERROMAGNETIC CRYSTALS, THEORETICAL ) /MA RAYS FROM ASSEMBLIES      13-006
                        SUPERCONDUCTIVITY IN THE CUAL2(C16) CRYSTAL CLASS.                                        15-062
FUNCTION AND RELATED NONCROSSING POLYGONS FOR THE SIMPLE- CUBE LATTICE.  HIGH-TEMPERATURE ISING PARTITION       02-067
CE IN RUBIDIUM- MANGANESE- IRON(3). DISCOVERY OF A SIMPLE CUBIC ANTIFERROMAGNET, ANTIFERROMAGNETIC RESONAN      06-038
                FERRO- AND ANTIFERROMAGNETISM IN A CUBIC CLUSTER OF SPINS.                                        02-065
ADOLINIUM ION.                                  CUBIC CRYSTAL FIELD SPLITTING OF THE TRIVALENT G           13-051
       THEORY OF NUCLEAR ACOUSTIC RESONANCE LINE SHAPE IN CUBIC CRYSTALS.                                         11-115
TTICE RELAXATION OF S-STATE IONS, DIVALENT MANGANESE IN A CUBIC ENVIRONMENT. ( THEORETICAL )       SPIN-LA      12-005
                    SPIN WAVE THEORY FOR CUBIC FERROMAGNETICS PART-3 MAGNETIZATION.                               02-011
```

Figure 4. Sample, Bell Laboratories Format

```
 68       have antiparasitic action on Entameba histolytica in rat [weanling]
 69        has antiparasitic action on Entameba histolytica in rat [weanling]
 70  not have antiparasitic action on Entameba histolytica in rat [weanling]
 71        has antiparasitic action on Entameba histolytica in rat [weanling] and weakly has toxi
 72        has antiparasitic action on Entameba histolytica in rat [weanling]
 73        has antiparasitic action on Entameba histolytica in rat [weanling] and has toxic actio
 74       have antiparasitic action on Entameba histolytica in rat [weanling] and have toxic acti
126        of amebic colitis caused by Entameba histolytica in rat [1275723,1275724, and 1275725
127        of amebic colitis caused by Entameba histolytica in rat [1275732 as dl(3- hydroxy-2- n
126        weakly have toxic action on Entameba histolytica in vitro and do not or weakly allevia
 25      THE LENGTH OF ACTION AND THE ENTERAL RESORPTION OF DIGITOXIGENIN- MONO DIGITOXOSIDE(Dt2
158   ouabain very strongly increases entry of calcium and strongly increases resting tension of
158       TENSION AND THE RATE OF NET ENTRY OF CALCIUM-45 IN ISOLATED PERFUSED RABBIT VENTRICLES
 20     of  hexobarbital by microsomal enzymes of liver
 21     of  hexobarbital by microsomal enzymes of liver
150 hydro ergokryptine with  dl hydro ergocornine and   dl hydro ergocristine[hydergine] inhibit
150  hydro ergocornine and   dl hydro ergocristine[hydergine] inhibit action of   vasopressin bu
150          tolazoline and  dl hydro ergokryptine with  dl hydro ergocornine and   dl hydro erg
  8       but, action antagonized by  ergotamine at higher dosage
  8   of rabbit, action increased by  ergotamine at low dosage but, action antagonized by  ergot
  5 dog and cat; action reversed by   ergotamine  ergotoxin   guanethidine and  phenylephrine bu
  5  action reversed by   ergotamine  ergotoxin   guanethidine and  phenylephrine but in rat, ac
119           Streptococcus pyogenes, Erslpelothrlx lnsldlosa and Streptococcus agalactae in vit
 42           in human accompanied by erythema [mild] given by injection into skin lesions
 46 pyrrolidone causes aggregation of erythrocytes in blood of hamster given intra-arterially
 46  A SYNTHETIC MACROMOLECULE ON THE ERYTHROCYTES OF THE BLOOD
162              chloramphenicol and  erythromycin strongly inhibit endotrophic sporulation of B
166                                   ERYTHROMYCIN- AND STREPTOMYCIN-LIKE ANTIBIOTICS AS BLEACHI
 32  acid in acid-soluble fraction of Escherichia coli
 36  amino uridine inhibits growth of Escherichia coli and Neurospora
 37       cytidine inhibit growth of Escherichia coli but do not inhibit growth of Neurospora
 35  amino uridine inhibits growth of Escherichia coli K-12, action reversed by  glutathione L-
 32 fluoro uracil has toxic action on Escherichia coli while organism is growing actively, actio
 34  amino uridine inhibits growth of Escherichia coli, action reversed by   glutathione ;, acti
 33   deoxy uridine inhibit growth of Escherichia coli, action reversed by  uridine   cytidine
119  growth of Staphylococcus aureus, Escherichia coli, Salmonella typhosa, Pasturella multocida
 32     pimelic acid in cell walls of Escherichia coli; increases content of N-acetyl hexos amin
 32   content of N-acetyl hexos amine esters and diamino pimelic acid in acid-soluble fraction o
157                                   estil general anesthetic decreases urinary output and incr
157      AFTER GENERAL ANESTHESIA WITH ESTIL.
 64       cyclo pentyl propionate and  estra di ol valerate hormone cause edema and thickening of
 53          THE TERATOGENIC ACTION OF ESTRADIOL AND THYROXINE ON MUELLER'S DUCT IN THE CHICKEN E
 63  increases excretion of  estrone  estradiol  estriol and total neutral 17- keto steroids in
 53                                   estradiol inhibits formation of Mueller's duct of chicken
 55                                   estradiol with  thyroxine strongly inhibit formation of st
 63 excretion of  estrone  estradiol  estriol and total neutral 17- keto steroids in urine of yo
 63                          URINE ESTROGEN RESPONSES TO HUMAN CHORIONIC GONADOTROPIN IN YOUN
136    methoxyestra-1,3,5-tri ene has estrogenic action
135 progestational action on immature estrogen-primed rabbit and does not inhibit growth of adre
131 progestational action on immature estrogen-primed rabbit given orally
132 progestational action on immature estrogen-primed rabbit given orally
133 progestational action on immature estrogen-primed rabbit given orally
143        action on immature rabbit [estrogen-primed] given subcutaneously
 63           increases excretion of  estrone  estradiol  estriol and total neutral 17- keto ste
  5      phenyl)-2-(iso propyl amino) ethanol have hypotensive action on barbiturate narcotized
144                 2-dl methyl amino ethanol increases incorporation of phosphorus into phospha
144  THE EFFECTS OF 2-DI METHYL AMINO ETHANOL ON BRAIN PHOSPHO LIPID METABOLISM
148           sulfate  choline phenyl ether bromide  DMPP and  hist amine acid phosphate in isol
133     sterone, less effective than  ethinyl testo sterone moderately have progestational actio
131           more effective than  ethinyl testo sterone strongly has progestational action o
132     sterone, equal in action to  ethinyl testo sterone strongly have progestational action
 43   ANTAGONISM OF LYSERGIC ACID DI ETHYL AMIDE BY CHLORPROMAZINE AND PHEN OXY BENZ AMINE.
 44  recognition of  lysergic acid dl ethyl amide in human if given simultaneously
 43  recognition of  lysergic acid dl ethyl amide in human only if given previous to latter
 13   catechol amines caused by  phen ethyl amine
 16   catechol amines caused by  phen ethyl amine
 14   catechol amines caused by  phen ethyl amine and does not inhibit secretion of catechol ami
 12   catechol amines caused by  phen ethyl amine  nicotine and   carbachol
 80    more effective than  α-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide has nicot
 80                        β-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide, more eff
113      substituted benzyl and phen ethyl hydr azines have toxic action [LD50 292,4000+, 4004,
145          tri ethyl tin and  tri ethyl lead very strongly inhibit metabolism of  glucose by
145  THE ACTION OF TRI ETHYL TIN, TRI ETHYL LEAD, ETHYL MERCURY AND OTHER INHIBITORS ON THE META
145 OF TRI ETHYL TIN, TRI ETHYL LEAD, ETHYL MERCURY AND OTHER INHIBITORS ON THE METABOLISM OF BR
146                                   ethyl mercury chloride  chlorpromazine  malonic acid [as s
145                              tri ethyl tin and  tri ethyl lead very strongly inhibit metabo
145                 THE ACTION OF TRI ETHYL TIN, TRI ETHYL LEAD, ETHYL MERCURY AND OTHER INHIBIT
 80       than  α-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide has nicotinic blockin
 80             β-3-(2-di ethyl amino ethyl) amino tropine bis meth iodide, more effective than
140     given as salts with di benzyl ethylene diamine
139                           α-9- ethyl-2'- hydroxy-2,5-di methyl-6,7- benzo morphan and  β-
139  morphan and  β- 5,9-di methyl-2- ethyl-2'- hydroxy-6,7-benzo morphan weakly have toxic acti
 83            1-(2-p- amino phenyl) ethyl-2- methyl-3- phenyl-3- propion oxy pyrrolidine has
```

Figure 5. Sample Page, Chemical Biological Activities

NON-IRRADIATED ABSORPTION OF D-GLUCOSE BY SEGMENTS OF INTESTINE FROM ACTIVE AND HIBERNATING, IRRADIATED AND NON-IRRADIATED GROUND SQUIRRELS, CITELLUS TRIDECEMLINEATUS NASA N63-11002(K) $2.60 0726

NON-ISOTHERMAL CORRELATIONS IN A NON-ISOTHERMAL PLASMA AD-290 053(K) $1.10 0196

NON-LINEAR INVESTIGATION OF MICROWAVE NON-LINEAR EFFECTS UTILIZING FERROMAGNETIC MATERIALS AD-290 572(K) $2.60 0487

NON-METALLIC BIBLIOGRAPHY AND TABULATION OF DAMPING PROPERTIES OF NON-METALLIC MATERIALS AD-289 856(K) $3.00 0502

NON-MILITARY NOTES ON NON-MILITARY MEASURES IN CONTROL OF INSURGENCY AD-290 237(K) $1.60 0696

NON-MOVING JUDGMENTS OF VISUAL VELOCITY AS A FUNCTION OF THE LENGTH OF OBSERVATION TIME OF MOVING OR NON-MOVING STIMULI PB 162 549(K) $1.60 0125

NON-RELATIVISTI TABLES OF NON-RELATIVISTIC ELECTRON TRAJECTORIES FOR FIELD EMISSION CATHODES AD-290 696(K) $14.50 0239

NON-SIMILAR NON-SIMILAR NUMERICAL METHODS OF SOLUTION FOR ELECTRODE BOUNDARY LAYERS IN A CROSSED FIELD ACCELERATOR AD-290 525(K) $5.60 0185

NONDESTRUCTIVE NONDESTRUCTIVE SYSTEM FOR INSPECTION OF FIBER GLASS-REINFORCED PLASTIC MISSILE CASES AD-289 825(K) $1.60 0632

NONDESTRUCTIVE X-RAY IMAGE SYSTEM FOR NONDESTRUCTIVE TESTING OF SOLID PROPELLANT MISSILE CASE WALLS AND WELDMENTS AD-289 821(K) $3.60 0637

NONDISSIPATIVE MAGNETOHYDRODYNAMIC STABILITY OF VORTEX FLOW - A NONDISSIPATIVE, INCOMPRESSIBLE ANALYSIS ORNL-TM-402(K) $3.60 0615

NONEQUILIBRIUM SCALE EFFECTS FOR NONEQUILIBRIUM CONVECTIVE HEAT TRANSFER WITH SIMULTANEOUS GAS PHASE AND SURFACE CHEMICAL REACTIONS. APPLICATION TO HYPERSONIC FLIGHT AT HIGH ALTITUDES AD-291 032(K) $1.60 0025

NONLINEAR APPLICATION OF VARIATIONAL EQUATION OF MOTION TO THE NONLINEAR VIBRATION ANALYSIS OF HOMOGENEOUS AND LAYERED PLATES AND SHELLS AD-289 868(K) $2.60 0667

NONLINEAR EXTENSIONS IN THE SYNTHESIS OF TIME OPTIMAL OR BANG-BANG NONLINEAR CONTROL SYSTEMS. PART I. THE SYNTHESIS OF QUASI-STATIONARY OPTIMUM NONLINEAR CONTROL SYSTEMS PB 162 547(K) $4.60 0235

NONLINEAR EXTENSIONS IN THE SYNTHESIS OF TIME OPTIMAL OR BANG-BANG NONLINEAR CONTROL SYSTEMS. PART I. THE SYNTHESIS OF QUASI-STATIONARY OPTIMUM NONLINEAR CONTROL SYSTEMS PB 162 547(K) $4.60 0235

NONLINEAR NONLINEAR FLEXURAL VIBRATIONS OF SANDWICH PLATES AD-289 871(K) $2.60 0669

NONLINEAR OPTIMUM NONLINEAR CONTROL FOR ARBITRARY DISTURBANCES NASA N62-15890(K) $2.60 0682

NONRECURRENT A TECHNIQUE FOR NARROW-BAND TELEMETRY OF NONRECURRENT PULSES AD-290 697(K) $2.60 0577

NONUNIFORM ELECTROMAGNETIC SCATTERING FROM A SPHERICAL NONUNIFORM MEDIUM. PART II. THE RADAR CROSS SECTION OF A FLARE AD-289 615(K) $2.60 0747

NONUNIFORM ELECTROMAGNETIC SCATTERING FROM ASPHERICAL NONUNIFORM MEDIUM. PART I. GENERAL THEORY AD-289 614(K) $2.60 0748

NORMAL PROBABILITY INTEGRALS OF MULTIVARIATE NORMAL AND MULTIVARIATE-T AD-290 746(K) $8.60 0760

NORMAL RESONANCE ABSORPTION OF GAMMA-RAYS IN NORMAL AND SUPERCONDUCTING TIN AD-289 844(K) $3.60 0826

NORMS NORMS FOR ARTIFICIAL LIGHTING AD-290 555(K) $1.10 0734

NORTH FACTORS INFLUENCING VASCULAR PLANT ZONATION IN NORTH CAROLINA SALTMARSHES AD-290 938(K) $7.60 0603

NORTH SONAR STUDIES OF THE DEEP SCATTERING LAYER IN THE NORTH PACIFIC PB 162 427(K) $2.60 0587

NORTH THE DEVELOPMENT OF RESCUE AND SURVIVAL TECHNIQUES IN THE NORTH AMERICAN ARCTIC PB 162 410(K) $12.00 0085

NOSE THE FLORA OF HEALTHY DOGS. I. BACTERIA AND FUNGI OF THE NOSE, THROAT, AND LOWER INTESTINE LF-2(K) $2.60 0458

NOZZLE FABRICATION OF PYROLYTIC GRAPHITE ROCKET NOZZLE COMPONENTS PB 162 371(K) $1.10 0351

NOZZLE FABRICATION OF PYROLYTIC GRAPHITE ROCKET NOZZLE COMPONENTS PB 162 370(K) $1.10 0353

NOZZLE FABRICATION OF PYROLYTIC GRAPHITE ROCKET NOZZLE COMPONENTS PB 162 372(K) $2.60 0352

NOZZLE THIRD SYMPOSIUM ON ADVANCED PROPULSION CONCEPTS SPONSORED BY UNITED STATES AIR FORCE OFFICE OF SCIENTIFIC RESEARCH AND THE GENERAL ELECTRIC COMPANY FLIGHT PROPULSION DIVISION CINCINNATI, OHIO OCTOBER 2-4, 1962. PLASMA FLOW IN A MAGNETIC ARC NOZZLE AD-290 082(K) $2.60 0147

NOZZLES HEAT TRANSFER AND PARTICLE TRAJECTORIES IN SOLID-ROCKET NOZZLES AD-289 681(K) $5.60 0030

NROTC DEVELOPMENT AND STANDARDIZATION OF FORMS 3 AND 4 OF THE NROTC CONTRACT STUDENT SELECTION TEST AD-290 784(K) $1.10 0201

NROTC EVALUATION OF NROTC AVIATION INDOCTRINATION FIELD TOURS FOR 1961-1962 AD-290 356(K) $1.60 0581

NUCLEAR A 7090 CODE FOR THE CALCULATION OF ELECTROMAGNETIC BLACKOUT FOLLOWING A HIGH ALTITUDE NUCLEAR DETONATION AD-291 141(K) $8.60 0372

NUCLEAR ACCURATE NUCLEAR FUEL BURNUP ANALYSES GEAP-4082(K) $1.60 0362

NUCLEAR APPLICATION OF NUCLEAR POWER SUPPLIES TO SPACE SYSTEMS TID-17306(K) $8.60 0741

NUCLEAR CAROLINAS-VIRGINIA NUCLEAR POWER ASSOCIATES, INC., RESEARCH AND DEVELOPMENT PROGRAM QUARTERLY PROGRESS REPORT FOR THE PERIOD APRIL - JUNE 1962 CVNA-156(K) $6.60 0839

NUCLEAR COMPUTER PROGRAMS FOR OPTIMUM START-UP OF NUCLEAR PROPULSION SYSTEMS TID-16730(K) $1.10 0712

NUCLEAR DOSE-TIME-DISTANCE CURVES FOR CLOSE-IN FALLOUT FOR LOW YIELD LAND-SURFACE NUCLEAR DETONATIONS PB 162 516(K) $1.60 0573

NUCLEAR EXTRUDED CERAMIC NUCLEAR FUEL DEVELOPMENT PROGRAM ACNP-62550(K) $4.60 0092

NUCLEAR FEASIBILITY DETERMINATION OF A NUCLEAR THERMIONIC SPACE POWER PLANT AD-290 068(K) $2.60 0031

NUCLEAR HIGH - ENERGY NUCLEAR PHYSICS RESEARCH PROGRAM AD-291 140(K) $1.60 0374

NUCLEAR HIGH-ENERGY NUCLEAR REACTIONS OF NIOBIUM WITH INCIDENT PROTONS AND HELIUM IONS UCRL-10461(K) $2.25 0222

NUCLEAR INVESTIGATIONS ON THE DIRECT CONVERSION OF NUCLEAR FISSION ENERGY TO ELECTRICAL ENERGY IN A PLASMA DIODE AD-290 727(K) $9.60 0385

NUCLEAR NUCLEAR SUPERHEAT DEVELOPMENT PROGRAM GNEC-254(K) $14.00 0386

NUCLEAR PRODUCTION OF TRITIUM BY CONTAINED NUCLEAR EXPLOSIONS IN SALT. I. LABORATORY STUDIES OF ISOTOPIC EXCHANGE OF TRITIUM IN THE HYDROGEN-WATER SYSTEM ORNL-3334(K) $.50 0617

NUCLEAR STRIKING EFFECT OF NUCLEAR EXPLOSION AD-290 824(K) $21.00 0083

NUCLEAR THE NUCLEAR PROPERTIES OF RHENIUM AD-291 180(K) $1.60 0310

NUCLEAR VARIATIONS IN THE TOTAL ELECTRON CONTENT OF THE IONOSPHERE AFTER THE HIGH ALTITUDE NUCLEAR EXPLOSION NASA N63-10486(K) $1.10 0142

NUCLEAR 630A MARITIME NUCLEAR STEAM GENERATOR GEMP-160(K) $8.10 0349

NULL-ZONE THE ESTIMATION PROBLEM IN NULL-ZONE RECEPTION FEEDBACK SYSTEMS AD-290 325(K) $11.00 0599

NUMBERS FUNDAMENTAL SOLUTION TO THE DIFFUSION BOUNDARY LAYER EQUATION FOR NEARLY SEPARATED FLOW OVER SOLID SURFACES AT VERY LARGE PRANDTL NUMBERS AD-291 031(K) $2.60 0023

NUMBERS LOCAL PRESSURE DISTRIBUTION ON A BLUNT DELTA WING FOR ANGLES OF ATTACK UP TO 35-DEGREES AT MACH NUMBERS OF 3.4 AND 4.7 NASA N63-10800(K) $.75 0516

NUMERICAL A MAINTENANCE PROGRAM FOR NUMERICAL CONTROL SYSTEMS ON MACHINE TOOLS TID-17376(K) $2.60 0809

NUMERICAL A PRIORI BOUNDS ON THE SISCRETIZATION ERROR IN THE NUMERICAL SOLUTION OF THE DIRICHLET PROBLEM AD-290 322(K) $4.60 0464

NUMERICAL NON-SIMILAR NUMERICAL METHODS OF SOLUTION FOR ELECTRODE BOUNDARY LAYERS IN A CROSSED FIELD ACCELERATOR AD-290 525(K) $5.60 0185

NYSTAGMUS MANIPULATION OF AROUSAL AND ITS EFFECTS ON HUMAN VESTIBULAR NYSTAGMUS INDUCED BY CALORIC IRRIGATION AND ANGULAR ACCELERATIONS AD-290 348(K) $1.60 0252

OAK A SAFETY REVIEW OF THE OAK RIDGE CRITICAL EXPERIMENTS FACILITY ORNL-TM-349(K) $5.60 0612

OBJECTS DRAG OF OBJECTS IN PARTICLE - LADEN AIR FLOW PHASE IV. BLUNT BODIES AND COMPRESSIBILITY EFFECTS AD-291 178(K) $6.60 0752

OBSERVATORY TONTO FOREST SEISMOLOGICAL OBSERVATORY AD-291 148(K) $3.60 0815

OCEAN A SAMPLE TEST EXPOSURE TO EXAMINE CORROSION AND FOULING OF EQUIPMENT INSTALLED IN THE DEEP OCEAN AD-291 049(K) $1.60 0582

OCEANOGRAPHIC OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI SEAS, SUMMER 1949. PART I SEA FLOOR STUDIES PB 162 426(K) $2.60 0585

OCEANOGRAPHIC OCEANOGRAPHIC AND UNDERWATER ACCOUSTICS RESEARCH AD-290 252(K) $2.60 0848

OCEANOGRAPHIC OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI SEAS, SUMMER 1949. PART IV. PHYSICAL OCEANOGRAPHIC STUDIES. VOL. I. DESCRIPTIVE REPORT PB 162 428-1(K) $3.60 0584

OCEANOGRAPHIC OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI SEAS, SUMMER 1949. PART IV. PHYSICAL OCEANOGRAPHIC STUDIES. VOL. I. DESCRIPTIVE REPORT PB 162 428-1(K) $3.60 0584

OCEANOGRAPHIC OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI SEAS, SUMMER 1949. PART IV PHYSICAL OCEANOGRAPHIC STUDIES. VOL. 2. DATA REPORT PB 162 428-2(K) $4.60 0586

OCEANOGRAPHIC OCEANOGRAPHIC CRUISE TO THE BERING AND CHUKCHI SEAS, SUMMER 1949. PART IV PHYSICAL OCEANOGRAPHIC STUDIES. VOL. 2. DATA REPORT PB 162 428-2(K) $4.60 0586

OCEANOGRAPHIC PROCEEDINGS OF INTERINDUSTRIAL OCEANOGRAPHIC SYMPOSIUM (NO. 1), BURBANK, CALIFORNIA, 5 JUNE 1962 PB 162 587(K) $2.60 0451

OCTYL RUBBER ELASTICITY IN HIGHLY CROSSLINKED SYSTEM

Figure 6. Sample, CEIR Format for Office of Technical Services

Figure 7. Sample Page, Diabetes Index

essentially the original Luhn format, and it should be noted in this connection that while Luhn recognized that the origin of the KWIC principle lay in the making of concordances, he claimed in particular the use of machines to achieve speed, completeness, and accuracy, and a novel format. 1/

The most common variant to the center position for the indexing window (or keyword position) is at the left or the beginning of the line. Netherwood's selected bibliography of logical machine design, which is probably the first of the modern permuted title indexes to appear in the open literature, used the left-most positions for the index entry word in each title listing. Slant marks were also printed to show the breaks in the normal order of the title (Netherwood, 1958 [437]). A proposed subscription service, advertized in 1958 but never actually brought into operation, would also have used the left-hand position. 2/

In these left position examples, the keyword-in-context principle is kept only partially intact since the word in the index position is directly adjacent to its most specific right-hand context, not to its left-hand. In variations such as developed at Stanford Research Institute, however, the index word is extracted from its context and printed separately in the left-hand margin, with the title in its normal order printed to the right. This type of variation has been called "KWOC", for keyword-out-of-context, and is illustrated in Figure 6, which shows the format developed by C.E.I.R., Inc. for the OTS index to U.S. Government Research Reports.

Table 1 lists a number of KWIC index projects for which computer programs are or might be made available to interested additional users. Computer programs have been written specifically for the IBM 650, 704, 1620, 709, 7090, and 7094 data processing systems, the G.E. 225 computer, the Deuce Computer in England, the UNIVAC 1103 and 1107 systems, and the Japanese computer JEIPAC, among others. In addition, some permuted title indexes are produced manually, or with the use of simple business office machine equipment. For example, an index to the AIBS Bulletin for 1951-1961 has been so produced by the American Institute of Biological Sciences. 3/

1/ Private communication, excerpt of letter from H. P. Luhn to C. L. Bernier, December 27, 1960: "With respect to the origin of the KWIC Index, you are, of course, right that it is a form of concordance, as stated in my original paper. Furthermore, keyword indexing has been practiced in various forms as far back as a hundred years ago. All of these methods were, however, dependent on manual effort. I would say that the significance of the present KWIC Index is based on the fact that it is produced automatically by machine, affording speed of compilation, accuracy and completeness. As far as the particular format of the Index is concerned, this is novel to my knowledge, in accordance with information I have been able to ascertain from others."

2/ "PILOT--a permutation index to this month's literature", see p. 8 and Figure 1. A left-most window full-title format was developed at Stanford University in cooperation with the IBM San Jose Laboratories. It has been applied by the Computation Center to the titles of computer programs for the benefit of users of the Program Library Computation Center, Stanford University, "The KWIC Index", 1963. See also Marckworth, 1961 [393].

3/ National Science Foundation's CR&D Report No. 11, [430], p. 10; Janaske, 1962 [299]; Shilling, 1963 [550] and [551].

Table 1. KWIC Type Indexes and Programs

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
Service Bureau Corporation - H. P. Luhn	"Bibliography and Auto-Index, Literature on Information Retrieval and Machine Translation"	First edition Sept. 1958 Second edition June 1959	2-column, 60-character single line title, center window	IBM 709	Basic Luhn KWIC
Chemical Abstracts Service	Chemical Titles	Semi-monthly	Standard Luhn IBM	1401	
Chemical Abstracts Service	Chemical Biological Activities	Bi-weekly-1st issue Sept. 1962	Single column Center window, 120-character line, upper and lower case, 120-character 1403 printer	1401	
Biological Abstracts	B. A. S. I. C.	Semi-monthly	Standard Luhn IBM	1440	Modified Luhn program: shading is used as an aid in scanning.
Biological Abstracts	Biochemical Title Index	Monthly	Luhn, Chem. Titles Formats	1440	
Bell Telephone Laboratories	-Index to the Literature of Magnetism -BTL talks and papers	Annually Annually	Single column, 120 character line, center window	7090	BE-PIP Program available through the SHARE organization
All-Union Inst. for Scientific and Technical Information			" . . . an index of the 'Chemical Titles' type."		Mikhailov, 1962 [418]

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
American Bar Foundation, Bobbs Merrill	Index to Current State Legislation	Initial issue, 1963			Eldridge and Dennis, 1962 [183]
American Diabetes Association	Diabetes-related Literature Index	First of proposed series, covering literature for 1960, issued 1963	2-column, left window, KWOC, full citation for each entry.	GE-225, Western Reserve program	
American Meteorological Society	Meteorological and Geoastrophysical Titles	April 1961, Oct. 1961, Jan. 1962 and following	Standard Luhn IBM	704	Includes a Systematic UDC-Subject Heading Index as well as modified KWIC.
Armour Research Foundation	Key words in context (reports received in document library)		Two column, 60-character line, center window	1103, 1107	
ASTIA (Defense Documentation Center)	Keywords-in-context title index. A list of titles for ASTIA documents not previously announced.	Irregularly No. 1, Oct. 1962 No. 2, Feb. 1963		IBM	
English Electric Company			KWIC-type	Deuce	See Black, 1962 [65]; Dowell and Marshall, 1962 [159].

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
General Electric Computer Dept. Phoenix	General Bibliography on Information Storage and Retrieval		Single column, center window	GE-225	
Gmelin Institute	Information Journal for Atomic Energy				See Koelewijn, 1962 [330].
Japan Information Center of Science and Technology				JEIPAC	"The JEIPAC, a transistorized information processing machine... has also been programmed for automatic indexing designed after the IBM KWIC indexing system." CR & D No. 11, [430], p. 120-121.
Lockheed Missiles and Space Division	KWIC Index of Reports		Modified Bell Labs.	1401/ 7090	See Carroll and Summit, 1962 [102].
Mimosa Frenk Foundation for Applied Neurochemistry	KWIC Index to Neurochemistry	August 1961	Standard Luhn IBM	IBM	
M. I. T.	KWIC Index to The Science Abstracts of China	1st Edition, December 1960	Standard Luhn IBM	704	

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
National Bureau of Standards	A Bibliography of Foreign Developments in Machine Translation and Information Processing	July 1963	Single column, 120-character line, center window	7090	Byproduct input from Flexowriter tape, citation data including upper and lower case, paper tape to punched card conversion. Walkowicz, 1963 [629].
National Bureau of Standards, W.W. Youden	-Index to the Communications of the ACM -Index to The Journal of the ACM		Single column, 120-character line, center window	7090	Youden, 1963 [659] and [660].
Radio Corporation of America	Significant Words Indexed From Title			RCA 301	Unpublished report by D. Climenson and M. Bechman
Stanford Univ. IBM San Jose Labs.	Dissertations in Physics	1961	Keyword-out-of-context, left window	IBM	Marckworth, 1961 [393].
Union Carbide Oak Ridge National Laboratory Libraries	Key Word Index Laboratory Reports Received Semi-annual Index January-June 1963	1st issue 1963, monthly thereafter	Bell Labs. System		
U.S. Atomic Energy Commission, Division of Technical Information	Index to Conferences Abstracted in Nuclear Science Abstracts	December 1963	Bell Labs. System		

Table 1. (cont.)

Issuing Organization and/or Investigator	Name of Index or Program	When Issued	Format	Computer	References and Remarks
University of California Lawrence Radiation Laboratories	Key-word-in-title (KWIT) index for reports	Various issues	Single column, 120-character line, center window	1401/7090	Records can be machine searched with and, or and not logic
University of California, Lawrence Radiation Laboratories	Unclassified Reports Titles List	Biweekly	Single column, 120-character line, center window	1401	By-product preparation from Flexowriter of library cards. Turner and Kennedy, 1961 [614].
University of Kansas	Kansas Slavic Index	Initial issue July 1963	60-character Modified Chemical Titles	1401	Farley, 1963 [192].
University of Kansas, University of Oklahoma	(Space Law collection)			1401	"Current research and development..." No. 11, p. 44 & 171.
Western Periodicals Company	Permuted Indexes to Scientific Symposia	As available	Standard Luhn IBM		Advertised regularly in various periodicals, e.g., Special Libraries

In addition to the regularly issued KWIC indexes by Biological Abstracts, Chemical Abstracts Service, the American Meteorological Society and others, a large number of special field, one time, or limited collection coverage indexes of this type have been and are being produced both in the United States and in other countries. Well-known examples include the programs developed at the Lawrence Radiation Laboratories, University of California, which simultaneously produce catalog, cross-reference and subject authority cards, [1] and the programs developed at the Bell Telephone Laboratories from 1959 onward (Kennedy, 1962 [310]).

Other KWIC indexing efforts cover a wide variety of subject matter. In the field of law, applications of KWIC type indexing include work on the legislation of the 50 states, a joint project of the American Bar Foundation and the Bobbs-Merrill Company (Eldridge and Dennis, 1962 [183], 1963 [182]), the ninth annual edition of the Index to Legal Theses and Research Projects, July 1962, (Eldridge and Dennis, 1963 [182]); and a cooperative program between the libraries of the Universities of Kansas and Oklahoma to prepare an index to the latter's "Space Law" collection. [2] In 1960, the KWIC Index to the Science Abstracts of China was prepared for an AAAS Symposium, (Henderson, 1961 [263]; Farley, 1963 [192]). At the University of Kansas Library also, the Kansas Slavic Index is being produced, with coverage of 3,000 articles from more than 200 Slavic journals. [3] In the computer technology field, Youden (1963 [659] and [660]) has compiled KWIC type indexes to both the Journal of the ACM and the Communications of the ACM and the Western Periodicals Company offers KWIC indexes to the proceedings of the Joint Computer Conferences as well as to the proceedings of other conferences and symposia including those in fields of electronics, aerospace and quality control. [4] A special-purpose application is in the use of a KWIC-index in lieu of cross-references in a revised edition of Current Medical Terminology. [5]

Examples of KWIC indexing projects abroad include work at the Japanese Information Center of Science and Technology, Tokyo, [6] an index "of the 'Chemical Titles' type" at the All-Union Institute for Scientific and Technical Information (VINITI) U.S.S.R., [7] an information journal for the atomic energy field being prepared at the Gmelin Institute, (Koelwijn, 1962 [330]), and work in Great Britain both at the English Electric Company [8] and the IBM British Laboratories (Black, 1962 [65]).

1/ Nation Science Foundation's CR&D Report, No. 11, [430], p. 42.

2/ Ibid, pp. 44 and 171.

3/ Ibid, p. 43; University of Kansas, 1963 [307].

4/ See advertisements in journals such as American Documentation.

5/ Gordon and Slowinski, 1963 [236], p. 55.

6/ National Science Foundation's CR&D Report, No. 11, [430], p. 120.

7/ Mikhailov, 1962 [418], p. 50.

8/ Dowell and Marshall, 1962 [159], p. 323; Black, 1962 [65], p. 316.

Trans-Canada Air Lines [1] is using a KWIC System, and at the EURATOM ISPRA laboratories a KWIC type program has been developed with up to 600-character context and a left-most indexing position. [2]

3.1.2 Advantages, Disadvantages and Operational Problems of KWIC Indexing

Luhn's original acronym, KWIC, is peculiarly apt for permuted title word indexing. As both proponents and critics have noted, the resulting product may be relatively crude in terms of indexing quality, but it is quick. The speed achievable both by elimination of human intellectual effort and by use of machine (especially computer) processing is indeed the major single advantage of this type of automatic indexing. Closely related, however, are the advantages of currency of announcement and the availability of these indexes for individual use.

Some typical claims with respect to speed and currency are as follows:

> "The permuted index was invented as a means of adequately controlling (essentially, of indexing) the literature without further intellectual effort, and thus eliminating indexing delays." [3]

> "The great merit of this particular method... is that it enables information concerning new articles to be made available very much more quickly than if there were the inevitable delays of human abstracting and indexing." [4]

> "In spite of the disadvantages which are pointed out, perhaps the greatest advantage is the timeliness and the speed with which permuted-title indexes can be prepared." [5]

Specific examples of high speed are given by Biological Abstracts, where one hour's computer time suffices to prepare and arrange entries for over 150,000 items. [6] Kennedy reports for the Bell Laboratories System that:

> "Editorial scanning is very fast; only several lines of print must be read for each report and the required text markings are trivially few. Keypunching, the largest single task, takes about two minutes per report... Main-frame time ...was 12 minutes for 1703 reports." [7]

1/ Simons, 1963 [556], p. 34.

2/ Meyer-Uhlenried and Lustig, 1963 [417], p. 229.

3/ Tukey, 1962 [611], p. 13.

4/ Cleverdon, 1961 [125], p. 108.

5/ Janaske, 1962 [299], p. 3.

6/ See Biological Abstracts, 36:24, p. xii.

7/ Kennedy, 1961 [311], p. 123.

Skaggs and Spangler claim:

> "The most obvious advantage of permuted indexing by computer is speed. In a test of one permuted indexing system, input of 3,000 punched cards containing titles and running text produced a permuted significant word index of 12,190 index entry lines, with approximately 85 minutes of computer time required for the permuting and sort operations. The output was printed at some 500 lines per minute..." 1/

In many cases, greater speed and timeliness are achieved at significantly lower cost. This is particularly true if the preparation of the input -- title, author, item identification and other descriptive cataloging information--serves multiple purposes from a single keystroking operation. Thus, the MATICO System provides from a single input (1) KWIC indexes as required, (2) selective dissemination notices to potential users of new acquisitions, (3) records on magnetic tape for the information retrieval file, and (4) book catalogs covering specialized areas of the collection, all at a net savings over previous methods of $0.39 for each title processed. 2/

Another advantage which is typically claimed for KWIC indexes is the use of the author's own terminology. The display of different words as they have been used in title context with any word looked up introduces "suggestiveness" so that different meanings and different browsing clues are shown. Kennedy makes the following typical points:

> "The use of the author's own terms--the alive currency of new ideas--rather than the considered reshapings to the indexing system may often be of advantage. The automatic generation as index entries of all the separate words in multi-term concepts is definitely so. Access is direct, under any one of the component terms, in the unrestricted manner of uniterm indexing. And context minimizes false drops; the author has supplied the term coordination." 3/

Others, however, consider some of these same factors to be definite disadvantages.

In general, even among enthusiasts of KWIC, there is more agreement as to the values of the technique as a device for current awareness scanning and as a dissemination index than for its use for more extensive searching. It was, in fact, primarily as a dissemination index that Luhn first proposed the KWIC technique. He pointed out that such indexes could be prepared with minimum effort and be ready for dissemination in the shortest possible time, justifying publication by inexpensive printing means. He also noted the following additional advantages:

1/ Skaggs and Spangler, 1963 [557], p. 30.

2/ Carroll and Summit, 1962 [102], p. 4.

3/ Kennedy, 1962 [310], p. 184.

"1. Because of the mechanical method of preparation, more information may be displayed than would have been practicable by conventional means.

"2. Keywords-in-context permit the cross-correlation of subjects to an extent not realizable by conventional procedures." 1/

The most common type of complaint against the KWIC indexing method is, as we have noted earlier, identical with that which is applied to word indexing in general--the lack of terminological control. Where the indexing terms are restricted to those used by the author himself, in his title or even full text, there arise many serious problems of synonyms, near-synonyms, homographs, neologisms, and eponyms. The effects of machine inability to resolve these problems are redundancy, scatter of references throughout the index, "haphazard groupings", 2/ and retrieval losses because the user is forced to guess at the terminology the author actually used. 3/ These problems are severely aggravated when only the title is used as the basis for index-word extraction.

Thus, a first and major question in attempting to appraise the effectiveness of KWIC-indexing techniques is that of the adequacy of titles alone as the source of subject content clues. Spurred on at least in part by the existence of KWIC-type indexes, several investigators have studied this question, with somewhat different results. Williams has explored for some years the possibilities of developing systematic procedures for title elaboration, especially making explicit information that is implied. Her conclusions are that indexing by title and direct elaboration of the title would produce index information equivalent to that found in Chemical Abstracts for about 50 percent of the documents studied, but that other procedures would be required for the remainder. 4/

Specific studies of title adequacy for a particular journal or field have been undertaken by both the American Institute of Physics and the Biological Sciences Communications Project. In the A.I.P. experiments, graduate physics students were asked to locate from limited clues certain specific articles appearing in The Physical Review, and search times were checked for their use of permuted title and other indexes. Another group of students compared the subject index entries in Physics Abstracts and Chemical Abstracts with the words in the titles of 25 papers from The Physical Review. In the case of Physics Abstracts, 69 percent of the entries for these papers were found in the words of the title and 63 percent of the titles contained all of the information supplied by the set of index entries. In the case of Chemical Abstracts, the corresponding percentages were 47 and 23. 5/ These latter findings, for the chemical index, are closely corroborated

---

1/ Luhn, 1959, [381], p. 295.

2/ Olney, 1963, [458], p. 44.

3/ See, for example, Dowell and Marshall, 1962, [159], p. 324: "This problem of 'conceptual scatter' becomes a nightmare when highly idiosyncratic author language is used as a basis for subject indexing."

4/ Williams, 1961 [643], pp. 361-363.

5/ Maizell, 1960 [392], p. 126.

Bernier and Crane who report that for the non-organic chemistry items covered by Chemical Abstracts, 34 percent of the entries can be derived from the titles. 1/

With respect to the Biological Sciences Communications Project studies, Shilling reports as follows:

> "Titles of scientific articles are being utilized at present in a great many ways under the general assumption that there is a positive correlation between the title and the content of the article. A study was undertaken to analyze the accuracy of titles in describing the content of biomedical articles. It was conducted in two parts. In part one, a group of scientists were asked to predict the content of selected scientific articles, in their area of interest, from the title, the author's name, and the name of the journal in which it appeared. The results of the first phase of the study on the first trial journal were so diverse as to make analysis impossible, and this part of the study was not pursued further. From this small segment of the study it appears that scientists are deluding themselves when they search by title only and then decide what they wish to read.
>
> "In the other half of this experiment, the article without title, author's name, or journal name was sent to 20 scientists, selected as experts in the scientific field of the article, who were asked to write a meaningful title. Fifty articles were used, five from each of ten selected biomedical journals. From this part of the study it is apparent that if the article is in a field which is relatively well standardized and has an accepted vocabulary, it is possible for a group of titlists to agree remarkably well on an appropriate title. However, if the article is loosely organized, contains more than one subject, or is in a specialty in which there is no standard vocabulary, then titling scientists fail to agree to a rather alarming extent." 2/

Other studies involving the question of usefulness of titles alone for indexing purposes include those of Doyle, Lane, Montgomery and Swanson, O'Connor, Ruhl, Swanson, and White and Walsh, among others. Doyle checked the retrieval loss likely to result from the synonymity-scatter problem for a permuted title index compiled in 1958 to the internal reports of the System Development Corporation. He found, for example, that for 12 direct references to McGuire Air Force Base, there were one to "New York Air Defense Sector", two to "New York Sector", ten to "NYADS" and five to "N. Y. Sector". 3/

1/ Bernier and Crane, 1962 [56], p. 120.

2/ Shilling, 1963 [551], pp. 205-206.

3/ Doyle, 1961 [166], p. 11.

Ruhl (1963 [506]) found that between 50 and 90 percent of author-prepared titles (the variation depending on subject field and other circumstances), did fully reflect the index terms assigned to these documents by human indexers. Lane and White and Walsh have also made studies directly related to the question of KWIC index effectiveness. The latter two investigators report only 52 percent retrieval effectiveness for a permuted title index to the Abstracts of Computer Literature, 1962, which they attribute to the changing terminology in the still new field of computer technology. 1/ Lane made counts of titles that would be "acceptable" and those that would not for a KWIC index for 50 titles drawn from each of 10 published indexes. He concluded that, if there were judicious pre-editing, technical articles in the technical subject indexes could be quite adequately covered, and papers in the fields of law, business, and the humanities somewhat less satisfactorily so, but that for the material indexed in the Reader's Guide to Periodical Literature, the KWIC technique would fail 58 percent of the time. 2/

Montgomery and Swanson have studied, as has O'Connor is even more detail, the adequacy of "machine-like indexing by people". Montgomery and Swanson took as their test corpus the September 1960 issue of Index Medicus and found that for 4,770 items, 85.8 percent contained either the word itself or a synonym for the subject heading assigned, slightly over 11 percent did not, and in the remaining cases the investigators could not clearly decide. They concluded, therefore, that: "Most of the articles studied could have been indexed by machine on the basis of machine 'inspection' of article titles alone." 3/ O'Connor, however, typically reports that of a random sample of 50 papers manually indexed under the term "Toxicity", five had titles which contained the word "toxic" or the word "toxicity" and 34 had titles which were not even indirectly connected with the term. ([443], [444], [445], [447] and [448]). With respect to the Montgomery-Swanson conclusions as such, Carlson raises the further critical questions of over-assignment and false drops and suggests that: "a simple machine processing of titles would give us way too much or practically nothing." 4/

Research activities at the American Bar Foundation have included checking of KWIC type indexing of several thousand legal articles with the subject headings assigned under the "Index to Legal Periodicals" system (Kraft, 1962 [333]). It is reported that:

1/ White and Walsh, 1963 [639], p. 346.

2/ Lane, 1964 [345], p. 46.

3/ Montgomery and Swanson, 1962 [421], p. 359. In another study (1962 [534], p. 468), Swanson reports findings for several thousand entries in classified bibliographies where approximately 90 percent of the sampled items contained title words that were identical, or similar in meaning, to the subject headings under which they were indexed. He notes, however, that similar results could have been produced by machine processing with the significant proviso that the machine have available an adequate synonym dictionary or thesaurus.

4/ G. Carlson, 1963 [100], pp.328-329.

"Interpretation of data revealed, among other things, that 64.4 percent of the title entries contained as keywords one or more of the ILP subject heading words under which they were indexed, and 25.1 percent contained logical equivalents. The remaining 10.5 percent of the title entries had non-descriptive titles." 1/

The difficulties with titles as sources of the indexing information stem from at least three distinct types of determining factors: (1) the language habits, background, interests, and idiosyncracies of the author; (2) the interests, familiarity with the subject matter, language habits, imagination, and idiosyncracies of the user, and (3) factors largely extrinsic to either the particular author or the particular user. In the first case, we find especially the problem of the witty, punning, deliberately non-informative title, the so-called "pathological title". Janske gives the provocative example, in the literature of information selection and retrieval itself, of "The Golden Retriever". 2/ Even in the non-pathological case, however, there is the serious question of whether the author himself is likely to be a good indexer. 3/

On the user side, the normal critical problems of "bringing the vocabulary of indexer and searcher into coincidence" (Bernier, 1953 [55]) are aggravated by the facts that the user of KWIC must anticipate the terminology used by a large number of different "indexers" (i.e., the authors), that title words spelled the same but with quite different meanings in different special applications are grouped together in the same place in the index, and that the same concepts may be expressed in quite different phraseology depending on the author's, rather than the user's, field of specialization. To these aggravating circumstances there must be added in turn the psychological acceptability to the individual user of the scatter and redundancy, to say nothing of the format and legibility, of a particular published index.

Such factors affecting the particular user will of course vary with the nature and purpost of his search. Kennedy points out, for example, that the location of a document from only a single clue, a single title word, is particularly easy with a permuted title index and he emphasizes that the "index purpose, use, size, statement and array are other factors of considerable moment in judging the value of title indexes". 4/

1/ National Science Foundation's CR&D Report No. 11, [430], p. 62.

2/ Janaske, 1962 [299], p. 4.

3/ See, for example, a report on a conference on better indexes for technical literature, ASLIB Proceedings, 13:4, April 1961, with a number of statements on the author as a poor indexer. See also Crane and Bernier, 1958 [144], p. 515: "Not even authors are qualified to index their own work unless they are equipped for the task by training and experience".

4/ Kennedy, 1961 [311], p. 125.

A major question in the area of user acceptability, however, is that of the adequacy of title alone to tell the searcher whether or not a specific document is relevant to his query or interest. A number of investigators, both documentalists and user-scientists, suggest that this is rarely the case. 1/ In fact, for many users, titles alone provide only a negative searching device--in an announcement bulletin or abstract journal the user's scanning of titles merely tells him whether or not he should read the abstract and then perhaps go on to the paper itself.

It is for reasons of this type, in all probability, that Montgomery and Swanson found less effectiveness of titles on relevance-judgment tests than might be suggested by their more optimistic findings as to the success of machine procedures for replicating human subject heading assignments. Whereas they have claimed that about 90 percent of test items could have been as successfully indexed by machine as by manual procedures, (Montgomery and Swanson, 1962 [421]; Swanson, 1962 [584]), they have also reported that: "Comparison of title relevance judgment with judgment based on full text examination indicates that titles are only about one-third effective (i.e., two-thirds of the relevant articles would be judged irrelevant) as the basis for estimating the relevance of the article to a given question". 2/ They go on to suggest, therefore, that "...indexing should be based on more than titles and...a bibliographic citation system should present to the requester something more than titles." 3/ Similarly, Jahoda reports in an analysis of 281 actual search requests at Esso Research and Engineering that only two-thirds could have been answered with a shallow index based on titles and major section headings of the documents and that answering the remainder of the requests would have required an index of considerable depth. 4/

The obvious factors affecting the utility of titles as the source of indexing-searching clues include, first, the limitation of most titles to the principal subject matter, the main topic or topics of the document. The display of title context does to some extent provide for modifications of the topic to the special aspects treated, but it is of course obvious that a title cannot possibly provide clues to subject content not implied in the words of that title. In many cases, the potential user wants information contained in the paper, or even

---

1/ See, for example, Atherton and Yovich, reporting on evaluations by physicists of experimental citation indexing, 1962 [26], p. 22: "The reliance on titles of papers for retrieval purposes was not sufficient"; Levery, 1963 [359], p. 235. "Titles are usually insufficient to furnish a correct index to the text"; Hocken, 1962 [274], p. 93: "The titles were not explicit enough"; Crane and Bernier, 1959 [145], p. 1053: "Lists of titles can be prepared rapidly, but they are inadequately useful in selecting articles of interest, and they provide little or no directly usable information"; Dowell and Marshall, 1962, [159], p. 324: "Frequently titles either lack sufficient detail or are in fact misleading"; Connolly, 1963 [136], p. 35: "Most titles are inadequate as descriptions of the contents of papers."

2/ Montgomery and Swanson, 1962 [421], p. 364.

3/ Ibid, p. 366.

4/ Jahoda, 1962 [298], p. 75.

in its appendices, which was not the principal concern of the author and may not even have been considered significant by him. The claim that the author, who knows his own subject best, has already indexed his paper best by his choice of words and emphasis in text, and especially in his title, is pertinent only to that main subject to which he addresses himself, not to the other potentially useful information which he may also disclose.

Other extrinsic factors affecting title adequacy and hence the effectiveness of title-indexes are the size and the relative homogeneity or heterogeneity of the collection or set of documents so indexed, the breadth or narrowness of the subject field or fields covered, the time period covered and whether for one or many fields. Whether or not material in more than one language is included is a special factor. These various factors interact in various ways, usually with disadvantageous effects when even the most "nondescript" human indexer (that is, one who accepts only words from the text itself) is replaced by "a keypunch operator whose job it is to convert the keywords into machine-readable form, and a machine whose job it is to assimilate machine-readable text and print out its permutations with each significant word serving as an access point." 1/

The difficulties of subject scatter, synonymy, homography, redundancy, and the like, however, will also occur in human indexing that relies heavily on title only, which is perhaps more frequently the case than is generally recognized, 2/ just as much as for machine-generated indexes involving the permutations of keywords in titles. Such disadvantages must therefore be balanced not only against the advantages of speed, timeliness, having an index announcement tool personally available at low cost, and the like, but also against the probability of obtaining as useful a tool within the limits of available human indexing resources and justifiable costs. Cleverdon, for example, comments as follows:

> "There are those who would say that this [KWIC] can in no way be called indexing, and that the value of such indexing must be very much lower than that done by intelligent trained human beings. This is a comfortable thought, but such small evidence as is at present available makes it appear doubtful as to whether it is entirely true. This is not to say that a human being cannot do a better job, but it certainly appears likely that the cost of employing a human being to do it is of doubtful economic value." 3/

1/ Herner, 1962 [266], p. 4.

2/ See, for example, Moss, 1962 [425], p. 39: "I am convinced that a great many of the UDC and other numbers which are provided on millions of cards in technical libraries up and down the country, and which look so erudite, are, in fact, no more than cards transliterating titles, with occasionally similar transliteration of a few randomly chosen words from the abstracts as well... We are, in effect, already largely using title indexing and complicating it unnecessarily by magic numbers." See also Crane and Bernier, 1958 [144], p. 514: "Some indexes to periodicals, particularly word indexes, are merely indexes of titles of papers or of abstracts."

3/ Cleverdon, 1961 [125], pp.107-108.

It is also of interest to note, moreover, that the very existence of machine-generated permuted title indexes should greatly increase the likelihood that authors will use better and more useful titles. 1/ At a seminar on word and vocabulary byproducts of permuted title indexing held at Biological Abstracts headquarters on October 8, 1962, Rigby of Meteorological and Geoastrophysical Abstracts reported informally that as of that time there was already discernible improvement in titles covered by their KWIC index. In the same year (1962), Tukey similarly stated that: "Chemical Titles has been heavily enough used to affect the construction of titles of papers on chemical subjects." 2/ Instructions to authors of the previously mentioned "Short Papers" 3/ for the A. D. I. 1963 Annual Meeting specified that at least six significant words should be included in their titles and nearly all authors did in fact comply. Two of the "Short Papers" are specifically directed to the topic of improvements that authors can make in writing their titles (Brandenberg, 1963 [80]; Kennedy, 1963 [312]).

Instructions of this type can be effectively used for situations where all authors are under the same administrative control, as in the internal reports prepared in a single organization. This type of situation, incidentally, is one for which KWIC proponents are often most enthusiastic (Kennedy, 1962 [310]; Black, 1962 [65]; Linder, 1960 [362]). Finally, there is considerable promise that pressures brought to bear by journal editors of the publications of professional societies, notably the American Institute of Chemical Engineers and other cooperating member societies of the Engineers Joint Council, will result in improved adequacy of titles and thereby increased effectiveness of title word indexes.

Certain other disadvantages of KWIC indexing techniques, however, relate specifically to operational problems and requirements in the machine production of these indexes. There is, first, the problem of the amount of context that is usually displayed--that is, the question of line length--and the related problems of title truncation and wrap-around. As Kennedy notes: "Progressive shifting of the title to bring a given word to the indexing column frequently causes portions of the title to exceed the line space available, first at the right margin, then the left, or even both simultaneously." 4/ A case in point is the perhaps apocryphal "EROTIC TENDENCIES AMONG TRAPPIST MONKS" where "ATHEROSCL" had been dropped off at the left.

For multi-column KWIC indexes, in particular, where the line length is typically 58-60 characters, "much of the relevance is lost because the reader sees the wrong slice of the title". 5/ The Bell Laboratories KWIC index, 6/ Chemical-Biological Activities, 7/

1/ See for example, Black, 1962 [65], p. 317; Youden, 1963 [658], p. 332.

2/ Tukey, 1962 [611], pp. 9-10.

3/ Luhn, 1963 [376] and [377].

4/ Kennedy, 1961 [311], p. 117.

5/ Brandenberg, 1963 [80], p. 57.

6/ Kennedy, 1961 [311], p. 118.

7/ Figures 4 and 5.

and Youden's indexes to ACM papers (1963 [659] and [660]) illustrate single-column formats that alleviate this problem by extending the title line to 103-106 characters, exclusive of the identification code. Youden has calculated that for the titles in the field of computer literature which he analyzed 30 percent of the titles would have been truncated in 60-character title line formats, but that only 2 percent would have been chopped by 103-character title length limits. 1/

A second disadvantageous effect of machine production requirements in most KWIC indexes is the tedious sequential scanning necessary because of the unbroken organization of the page format and the long blocks that occur for frequently occurring word entries. Doyle (1959 [168], 1961 [166]) has investigated this problem of block length and suggests either that alphabetization be carried out to the words following those in the indexing window or that the entries in the block be permuted also in a second-order cycle. The latter suggestion has the advantage of facilitating any two-term coordinate indexing type of search, "because one can now look up directly any pair of subject words, regardless of whether or not they occur adjacently in a sentence." 2/

Redundancy in KWIC indexes, which aggravates the sequential scanning and the long-block fatigue effects, is in large part the result of difficulties in establishing the most appropriate bounds for exclusion or "stop" lists. We have previously distinguished machine-generated indexes of the derivative type from certain of the machine-compiled indexes primarily on the basis that in the first case, the criteria for determining the significance of the keywords to be used as the index access points are applied automatically during the machine processing, even if the selectivity so achieved is only "negative selectivity." 3/ The amount of index entry redundancy, of too many entries and of irrelevant entries is, in simple KWIC indexing, a direct function of the length and contents of the stop list.

In Luhn's original proposals for both KWIC and other types of automatic indexing, he pointed out the importance of the rules which must be established in order to differentiate the significant words from the nonsignificant. He says, for example:

> "Since significance is difficult to predict, it is more practicable to isolate it by rejecting all obviously nonsignificant or 'common' words, with the risk of admitting certain words of questionable value. Such words may subsequently be eliminated or tolerated as 'noise'. A list of non-significant words would include articles, conjunctions, prepositions, auxiliary verbs, certain adjectives, and words such as 'report','analysis', 'theory', and the like." 4/

1/ W. W. Youden, 1963 [458], p. 331.

2/ Doyle, 1961 [166], p. 13.

3/ Artandi, 1963 [20], p. 15.

4/ Luhn, 1959 [381], p. 289.

Interesting variations are to be noted in the current practices of using stop lists. Some lists are quite short, and others extend to several thousand words. Parkins reports that a mere 14 words on the stop lists used for B. A. S. I. C. are responsible for 80 percent of the title lines that need not be printed, but that their original list of 200 stop words grew quite rapidly to more than 1,000 now in use. 1/ Chemical Abstracts Service representatives reported in 1962 an initial list of about 1,000 words which dropped to 300 at one time and then was increased again to the original level. 2/ Using a stop list of 82 words eliminated 30 percent of a 42,000-word corpus of internal reports at the System Development Corporation, (Olney, 1961 [456]).

Critical questions in the establishment of stop lists relate to the problem of balancing the economics of the number of title lines to be printed and to be subsequently scanned against the loss of retrieval effectiveness if certain words are omitted from the search entry positions. How this balance should be achieved may vary from one subject field to another and between different organizations. In several regularly published KWIC indexes, the actual list used to exclude the presumably nonsignificant words is printed so that the user can check before proceeding to actual search. Williams has suggested that each excluded word be listed once, in its proper alphabetic place in the index, if it occurs in the titles of the particular set of items being indexed. 3/

In general, however, not enough is yet known about the requirements of particular subject fields and particular types of organization to arrive at the most effective compromises in establishing exclusion lists for keyword indexing. Noting that stop lists in actual use vary from only a few function words such as prepositions and conjunctions to lists several hundred words long, Brandenberg points out that:

> "At the present state of the KWIC indexing art the selection of stop words appears to be largely arbitrary and a comparison of half a dozen stop lists shows that they have about two dozen words in common." 4/

Kennedy and Doyle both specifically suggest that more research on the contents and effects of stop lists is necessary, (Kennedy, 1961 [311], 1962 [310]; Doyle, 1963 [162]), but Kennedy points out the ease with which the machine programs themselves can be used for modification of the lists. 5/

1/ Parkins, 1963 [466], p. 27.

2/ F. A. Tate, discussions at seminar on the word and vocabulary byproducts of permuted title indexing, Biological Abstracts headquarters, October 8, 1962.

3/ T. M. Williams, discussions at seminar on word and vocabulary byproducts of permuted title indexing, Biological Abstracts headquarters, October 8, 1962.

4/ Brandenberg, 1963 [80], p. 57.

5/ See also Clark, (1960 [123], p. 459), who suggests: "It is very probable... that the cut-off points [for most common, for very infrequent, words] will have to be adjusted to the material we actually use. The effect on the process of such factors as style, size of text, the complexity of the subject matter, and the like, is as yet not clearly seen. The collection of large amounts of text and their analysis will undoubtedly be the best way of determiningthe effects of these variables."

Some of the reasons for keeping stop lists short, however, may reflect unnecessary programming difficulties. Turner and Kennedy have reported that in the SAPIR system a title word is compared only with the group of nonsignificant words that have the same number of characters, in order to reduce the machine time required for the exclusion list search. 1/ Skaggs and Spangler give an account of an exclusion list system developed for general text processing as follows:

> "A representative form developed by General Electric is composed of three groups of words, high frequency, special and standard. The high frequency words (25) occur most frequently in English text. A compression of approximately 35 percent will occur for most kinds of text when these 25 words are deleted. The special words are derived from the particular body of text being processed. The composition of this group is left to the program user. Normally the words for this group are selected by making an Editing list in alphabetical sequence. The words appearing in the index position on the preliminary listing are then reviewed.
>
> "Standard words are words that occur with a relatively high frequency in most types of text and therefore are appropriate for a general purpose screen. In the GE program, 375 words are used in this group.
>
> "To minimize computer processing time, it is desirable that words in the Exclusion Dictionary be arranged in approximate order of their frequency of occurrence." 2/

It should be noted, however, that in most cases stop list searches can be programmed in the form of so-called "logarithmic", "partitioning" or "bifurcation" searches in which the number of machine operations required is only $\log_2 N + 1$, where N is the number of words in the list.

The more words excluded, the fewer the title entry lines that must be included in the final index. This is a factor involving first of all the user in the sequential scanning he must do, where, as Coates has remarked, the retrieval effectiveness is usually in inverse proportion to the amount of such scanning required. 3/ Secondly, longer stop lists help to minimize the long block problem, since it is obviously the most frequently occurring title words that have not been excluded that cause the longest blocks of entries.

1/ Turner and Kennedy, 1961 [614], p. 7.

2/ Skaggs and Spangler, 1963 [557], p. 29.

3/ Coates, 1962 [134], p. 430.

The important economic factor, however, is the total number of lines to be printed in the index, which is directly reflected in page costs. The effects of page costs, in turn, engender compromises in printing quality, such as page format and size of type. These are among the serious unresolved problems that affect user acceptance of KWIC indexes and involve questions of format, legibility, character sets, and size of the index.

In general, however, in the present state of the art of KWIC indexing, the consensus seems to be that of qualified praise, especially for the early announcement and dissemination applications. The KWIC index is recognized as responding to a definite need,[1] as having merit for fields in which more conventional indexes do not exist as well as for current awareness searching,[2] as receiving excellent response from users "because they can take a handy booklet, sit down at a table and look under the words they know and use, and which they expect other engineers to use in titles."[3] Bernier and Crane, after considering comparative effectiveness data for subject as against word indexing, come to the following conclusions:

> "Title lists keyed by words have value for quick distribution and fast use since time is often a very important element in the obtaining of information. Such lists do not serve adequately for thorough searching. ... A title concordance may be more useful than would seem from the ... data on index entries. However, it must obviously be incomplete, must have many unnecessary entries, and would not prove suggestive enough to users who lack background in the subjects sought."[4]

Additional benefits can quite readily be obtained by taking advantage of the bibliographic information once it is in machine-readable form to provide selective KWIC indexes (Balz and Stanwood, 1963 [28]; Black, 1962 [65]; Carroll and Summit, 1962 [102]) machine retrieval of item citations by specified keywords. (Kennedy 1961 [311]) and selections of items geared to a Selective Dissemination of Information System (Barnes and Resnick, 1963 [36]; Balz and Stanwood, 1963 [28]). Gallianza and Kennedy at the Lawrence Radiation Laboratory, for example, report as being under development programs for the IBM 1401 and 7090 computers which will combine KWIC type indexing features with the logical search operators "AND", "OR", and "IF" in order that users may specify subject searches in ordinary English language terms.[5]

1/ Clapp, 1963 [122], p. 7.

2/ Markus, 1962 [394], p. 19.

3/ Black, 1962 [65], p. 316.

4/ Bernier and Crane, 1962 [56], p. 120.

5/ National Science Foundation's CR&D Report No. 11, [430], p. 42.

3.2 Modified Derivative Indexing

Some of the more obvious of the disadvantages of KWIC indexing techniques can be reduced if not eliminated by a variety of human and machine procedures. These include augmentation of titles to provide additional clues to subject aspects, manual post-editing, and synonym reduction through such devices as thesaurus lookups.

The ink was scarcely dry on the first issues of a KWIC index before a number of suggestions for improvements, modifications, and augmentations were proffered in the literature. In fact, both Luhn and Baxendale considered various possible refinements in their original proposals. The first systematic review of work in the field of automatic extracting--whether to produce indexes or abstracts, or both--was made by Edmundson and Wyllys in 1961 [181]. They covered not only the KWIC type indexes as such, but also modifications suggested by Baxendale, Luhn, Oswald and others, and they themselves advanced a number of additional possibilities. Of the various modifications and refinements that have been suggested, the most obvious is that of title augmentation.

3.2.1 Title Augmentation

The machine-prepared index that was probably the first to go into productive operation is actually one involving title and subject indicators rather than pure keyword-from-title permutations. The CIA project, beginning in 1952, is based upon manual pre-editing of the titles themselves, with the words to be picked up as index entries being underlined. In addition, it involves assignment of other words, descriptors or terms from a hierarchical classification schedule to indicate additional access points (Veilleux, 1961 [624].

In later KWIC type indexing, the possibilities of improving effectiveness by pre-editing or post-editing to modify and expand titles have been suggested and explored by a number of investigators. The semi-automatic indexing reported by Janaske adds descriptive words or phrases in parentheses at the end of titles and uses them as additional indexing points (Janaske, 1962 [299]). At Biological Abstracts Service, improvements have been obtained (without sacrifice in the speed desired in order to index 5,000 abstracts twice a month) by title supplementation as well as by an improved stop list and by post-editing word divisions and word recombinations. 1/ Titles for each of two 12,000-item bibliographies in the field of radiobiology are reported as being edited considerably before KWIC type processing. 2/ Other examples of modified derivative indexing based on title augmentation include Chemical Patents 3/, the Applied Physics Letters indexing project at Oak Ridge National Laboratory, which provides for an author-prepared form to describe features of property and method not covered in the title, 4/ and the KWIC Index to Neurochemistry ([420]).

1/ Parkins, 1963, [466], p. 27.

2/ Davis, 1963 [150], p. 238.

3/ See Markus, 1962 [394], p. 19, and ref. [662].

4/ Connolly, 1963 [136], p. 35.

To some extent, however, the use of human editors to improve the product of KWIC type indexing defeats the initial purpose of a quick and purely clerical or mechanical process. Thus, Dowell and Marshall argue:

> "... The basic permuted-title index can be substantially improved by editing and re-writing the titles before they are submitted to the computer. ... But this of course, destroys the great advantage claimed for the permuted title index, 'that it is a purely clerical process'. Intellectual effort has entered the picture again and we are back where we started." 1/

In the extreme case, the re-introduction of intellectual effort is in effect the re-introduction of conventional human indexing, with the machine's role limited to that of compilation, as in the case of the "notation-of-content" statements prepared for NASA's STAR System (Slamecka and Zunde, 1963 [561]; Newbaker and Savage, 1963 [430]).

Kennedy suggests instead, therefore, that the augmentation might be accomplished by the authors themselves. However, it may then be pointed out, as by Bernier and Crane, for example, that the supplementation of titles before publication in order to provide suitable additional indexing words would be "awkward, space-consuming and difficult". They continue:

> "It would call for the attention of index experts at the manuscript stage, which would delay publication and expand the total indexing effort. Furthermore, good, thorough indexes are based on the full information of abstracts and papers, not on their titles only." 2/

An alternative method for title augmentation to improve the quality of KWIC indexing is therefore to establish procedures for machine selection of significant words from more of the text than just the titles alone. In fact, Luhn himself did not limit his technique as originally proposed to titles only but indicated that the process could be performed at various levels: title, abstract, or full text. 3/ In the 1958 permuted index to the ICSI preprints, entries were derived from titles, author's names, author affiliations, headings within the paper, figure and table captions, and sentences and phrases taken directly from text. 4/ Combinations of human and machine procedures based on sentences and phrases selected from text are described by Herner who cites a two-fold advantage: "First, it is not wholly dependent on the informativeness or lack of informativeness of titles and bibliographic citations, and, second, it affords a greater depth of analysis than is generally possible where titles or bibliographic descriptions alone are used." 5/

1/ Dowell and Marshall, 1962 [159], p. 324-325.

2/ Bernier and Crane, 1962 [56], p. 117.

3/ Luhn 1959 [381], p. 289.

4/ Citron, et al, 1958 [120], p. i.

5/ He rner, 1963 [264], pp. 1-2.

Taking more text as the basis for automatic derivative indexing adds, of course, the problems and costs of keystroking additional input material. At the same time, most of the major problems of scatter of references, synonymity, redundancy and exclusive reliance on the author's own language and terminology not only remain but may quite probably be intensified. The problems of establishing suitatle rules for selection of significant words are aggravated, not only by the far larger number of different words to be processed, but because of unresolved problems in effectively relating length of index and depth of indexing to the length of the document. 1/

There are, however, a number of practical suggestions by which machine augmentation of titles might be accomplished. First is the invariant selection of words that are capitalized, other than those that begin a sentence. 2/ As Wyllys points out, this type of selection criterion would emphasize proper names, and these in turn might be particularly valuable clues, especially in a military intelligence situation. 3/ It has also been suggested that the selection criteria should depend on particular pre-specified contexts, such as being preceded by the words: "the results were...,", "in conclusion ...", and the like.

A second type of machine selection procedure is the converse of the exclusion or stop list, namely, an inclusion list or dictionary which may involve especially significant words for a particular subject matter area or words that are of importance to a particular organization. In the discussions of the Area 5 ICSI papers it was remarked:

> "Another complication is that mechanized indexing finds in a paper what was important to the author. What happens if there is something in the paper not important to the author but of importance to the indexer? One possibility is to have a list of words and phrases expressing the interests of a particular collection, which the machine looks for in the papers. If this word or phrase occurs even once, it should be picked up as an indexing term." 4/

1/ See, for example, Wyllys, 1963 [653], p. 22.

2/ See Luhn, 1959 [371], p. 52; [384], p. 8.

3/ Wyllys, 1963 [653], p. 15.

4/ See Ref. [578], p. 1263. See also, among others, Luhn, 1959 [371], p. 52: "Just as common words have been eliminated by look-up in a special index, certain essential words may be looked up in another special index for the purpose of listing them under any circumstances".

This approach to the selection problem can be combined with other devices, as in the "Selective Dissemination" system described by Kraft in which keyword extraction indexing is applied to abstract, title, author's name and manually assigned index terms, after processing of all input material against both "in" and "out" dictionary lists. 1/

The use of abstracts rather than full text as source material makes the selection criteria problems somewhat less severe. In addition, there is evidence to suggest that the abstract does contain much of the significant information that would normally be indexed and the text of the abstract is therefore a fertile field for title augmentation. In experiments conducted by Slamecka and Zunde on the comparison of indexing terms manually assigned with the occurrences of the names of these terms in abstracts used in NASA's STAR system, it was found that 80.4 percent of the assigned terms were contained in the abstracts. 2/ Swanson, on the other hand, suggests that, at least for short articles having homogeneous subject matter, title and first paragraph "are nearly as good as full text." 3/

A combination inclusion-exclusion list system may involve prior "weighting for relevance "of words that are judged by human analysts to be significant for purposes of search and retrieval, as suggested by Swanson, for example:

> "The computer first separates those words which are important for purposes of information retrieval from those which are unimportant. This is accomplished by means of looking up each word in an alphabetized word list with which the computer is furnished. Each word in this word list carries a 'weight' which reflects an estimate of its importance for retrieval purposes. Words of zero weight are completely unimportant and discarded by the computer for indexing entries." 4/

Continuing work at Thompson Ramo-Wooldridge on automatic indexing methods includes further investigation of assignments of relevance weight estimates to words and phrases, (1959 [490] and [491], 1963 [602]).

3.2.2 Book Indexing By Computer

For internal indexing, that is, the subject indexing of the contents of a single book or report, automatic indexing experiments are usually directed toward the processing of full text, with use of stop lists of various lengths. The work of Artandi for her doctorate

---

1/ Kraft, 1963 [334], pp.69-70.

2/ Slamecka and Zunde, 1963 [561]. In addition they report (p. 139) that a large number of the terms not found were "either broad, general terms (i. e., 'device') or generic level concepts of terms contained in the abstracts."

3/ Swanson, 1963 [580], p. 1.

4/ Ibid, p. 1.

at Rutgers in indexing of a book by computer programs (1963 [20] and [22]) is an example of such modified derivative indexing. Specifically, Artandi's method involves:

(1) Establishment of a list of key terms appropriate to a given subject area to be used as an inclusion list for word extractions from text.

(2) Application of an appropriate syndetic apparatus to be used in the compilation and ordering of the index entries.

(3) Means for the automatic selection of index entries other than those on the pre-specified inclusion list, especially for the selection of proper names.

The text used by Artandi for her study consisted of a 59-page chapter on halogens from J. W. Mellor's Modern Inorganic Chemistry. This text was keypunched with special tags being assigned to indicate the page numbers and the incidence of capitalized words in the text. Text words greater than three characters in length were first checked against the inclusion dictionary of "detection terms". There was, in addition, an "expression term" dictionary which constituted the vocabulary of the final index and in which a given expression term might or might not be identical with the corresponding detection term. Cross-references were supplied by a program routine which checks the index term list against a list of expression terms with their detection terms grouped under them and which compiles cross-reference entries, one for each detection term associated with an expression term appearing on the index list.

For her experimental corpus, Artandi's program developed 363 page references, 138 different index entries and 35 cross-references. She compared these results with those obtainable by conventional human indexing with respect to the factors of heading density (ratio of number of entries to number of words in the book), entry density (ratio of the number of page references to the number of pages), and distribution (ratios of entries for chemical compounds, proper names, and subject entries to the total number of entries. No indexing errors were found in the computer-generated index for a 5 percent random sample of the pages of the corpus, but five omissions were found in the machine indexing of these sample pages. Artandi concluded, however, that although the quality of indexing appeared favorable, the costs, which approximated $1.50 per page indexed, were impractically high.

Book indexing by computer has also been investigated by Maloney, Dukes, and Green at the Army Biological Laboratories, Fort Detrick, Maryland.[1/] Input is based on the by-product paper tape generated when the manuscript is typed on a tape typewriter. The paper tape is in turn converted to punched cards which are then processed by a UNIVAC SS-90 II computer in an editing run that deletes unrecognizable codes and then stores page,

---

1/ C. J. Maloney, private communication. A report by C. J. Maloney, J. Dukes, and S. Green, "Indexing reports by computer" is in process of preparation for publication.

line, sentence number and other reference identifications. After re-processing against a stop list of common words, all other words in the edited text are selected as candidate index entries, these are then sorted into alphabetical order with subsequent printout giving each word occurrence followed by the entire sentence which contained it and the page and other location identifications. This computer output is then post-edited manually not only to eliminate trivial entries but also to normalize terms and phrases used.

3.2.3 Modified Derivative Indexing - Baxendale's Experiments

As has been previously noted in the introduction to this report, the name of Phyllis Baxendale together with that of H. P. Luhn is generally accorded credit for pioneering efforts in the entire area of automatic indexing. Baxendale in particular is generally credited with the first actual experiments in modified derivative indexing. In investigation beginning in the late 1950's, she has explored not only statistical approaches to automatic selection of index terms (based for example on word frequencies) but also the use of word pairs, word groups, contextual associations, and in particular the subject-indicating clues of prepositional phrases (Baxendale, 1958 [41], 1961 [40], 1962 [42]; Becker, 1960 [44]; Edmundson and Wyllys, 1961 [181]).

Baxendale began by considering the patterns of scanning that humans typically use to select "topic" sentences, phrases and words, and she then proceeded to simulate by computer program the selection of phrases consisting primarily of nouns and modifiers. In her first experiments, (1958 [41]) she used two methods of automatic selection. In the first procedure, words serving the grammatical functions of pronoun, article, auxiliary verb, conjunction and the like, were deleted by stop list lookup. Frequency count statistics were then derived for the remaining words. In her second procedure, the computer was programmed to select prepositional phrases from text and to use the four words succeeding the preposition as index entries unless an additional preposition or a punctuation mark is first encountered.

In later experiments, Baxendale has explored possible grammatical models "which would select all and only nouns or adjective-noun combinations". 1/ Taking as an initial corpus a sample of document titles, rules were devised to reject for human analysis titles with question-marks and the like, to eliminate numeric information and single symbols, and to segment the title into its component clauses and phrases by the detection of commas, periods, and similar clues. By list lookup, certain words are identified as capable of serving the syntactic functions of being quantifiers, prepositions, or clause introducers. Special subscripts are then assigned to these words and the subscripts are examined by machine to provide further segmentation; to delete quantifiers, auxiliary verbs, or words ending in "ed" or "ing" and preceded by an auxiliary verb, and to determine relationship functions between the remaining, presumably substantive, words.

Still other work by Baxendale has been directed toward the development of frequency of co-occurrence or textual association of candidate indexing terms. She reports as follows:

1/ Baxendale, 1961 [40], p. 209.

"[In the frequency matrix]... the diagonal elements ... give the total frequency of an index term and the off-diagonal gives the frequency of co-occurrence of two terms. The diagonal of the 'context' matrix represents that portion of the total vocabulary with which an individual term has been coordinated, and the off-diagonal the extent to which two terms have common context... Such matrices give a basis for examining the extent to which terms are generic or specific within the context of the collection of documents. One can speculate that terms occurring with high frequency and wide context, i.e., with frequencies distributed amongst all or nearly all off-diagonal elements of the matrix are of such broad connotation as to be indifferent discriminators of content ... The frequency and context matrices can again be used to determine the modifiers with which they can most meaningfully be coupled for the collection of documents being considered." 1/

Finally, Baxendale notes that on the basis of her studies it should be possible to select quasi-subject headings based on frequency counting criteria, but then to order the remaining vocabulary of selected terms according to contextual measures of association which are semantic, syntactic, or statistical in nature. Experimental results for a collection of 1,500 documents included semantic associations between "searching" and "retrieval", syntactic associations of "machine" or "literature" with "retrieval", and the apparently misleading association of "metal" with "retrieval" which, however, had statistical significance within the particular document sample. 2/

Other investigators who have explored noun-adjective clues for selection include Anger, Chonez, Langleben and Shumilina, and Swanson. Anger looked for relationships indicated by syntactic dependencies or by noun-adjective and adjective-adverb linkages, and gave in an appendix a suggested program for phrase inversions. 3/ Chonez has described a computer program which by recognizing "separating" words, especially prepositions, and applying "pseudo-grammatical" rules compiles an index to English language items in the fields of ionized gas physics and thermonuclear fusion. It is claimed that:

"The subject index thus prepared is similar in presentation to Luhn's KWIC indexes, but is fundamentally different in conception and is in fact intermediate between... (this) ... and the conventional alphabetic subject indexes." 4/

Langleben and Shumilina are concerned with machine-aided procedures for translation from natural language materials to an intermediary or documentation language.

1/ Ibid, pp.215-216.

2/ Ibid, pp. 216-217.

3/ Anger, 1961 [15], pp. III-6 ff.

4/ Chonez, et al, 1963 [119], p. 31.

They indicate, for example, that the preposition "from" serves as a key for the treatment of two nouns connected by it. [1] Swanson, describing research project progress at Ramo Wooldridge as of 1960, reported to the National Symposium on Machine Translation with respect to multiple meaning problems as follows:

> "We are also investigating the possibility of discovering semantic attributes of words based upon certain automatically recognizable statistical features of the context. Our initial endeavor in this direction has been to attempt to discover a classification system for nouns based upon their frequency spectrum of categories of modifying adjectives, these categories being automatically recognizable." [2]

3.3 Derivative Indexing From Automatic Abstracting Techniques

While Baxendale's work has had certain points in common with automatic abstracting or extracting processes, particularly in the use of word frequency statistics and the consideration of possibilities for first selecting topic sentences, her major interests in this area have been in automatic indexing as such, rather than in machine selection of sentences from text to serve as an automatic extract or derivative abstract of the document. Much of the machine processing to date of full text for documentation purposes, however, has had the latter goal as the principal research objective.

As we have previously noted, the subject of automatic abstracting or auto-condensation is not in itself a primary concern of this survey. Nevertheless, the significant words occurring in the abstract of a document, whether generated by man or by machine, are obviously good candidates for indexing terms. Moreover, it has been strongly suggested that the questions of using positional, editorial, and syntactical clues in order to improve automatic indexing techniques will profit by research that is being done in both automatic extracting procedures and in other types of linguistic data processing based upon full text. [3]

3.3.1 Auto-Condensation and Auto-Encoding Techniques of H. P. Luhn

Although Luhn's work in the field of documentation aided by machine has had its best known and most popular acceptance with respect to the KWIC index proper, even more provocative possibilities lie in the development of some of the auto-condensation and auto-encoding techniques which he also proposed, especially for full text processing. In this area, although he himself has also suggested a variety of possible improvements and refinements, the actual experimental work done by him and by his associates has mostly been done on the basis of word frequency statistics.

---

[1] Langleben and Shumilina, 1962 [347], p. 109.

[2] Swanson, 1961 [585], pp. 391-392.

[3] See, for example, Wyllys, 1963 [653], p. 7.

Considering first the most frequently occurring words in a given text as too common to be subject-indicative (those usually stopped or purged by a suitable exclusion dictionary or stop list, for example) and next the least frequent words as being rarely topical in a content-revealing sense, Luhn settles upon a middle range of frequency of word occurrence as the basis for his auto-condensation processes. The actual frequency counts are computed, together with indications of page, line, and occurrence within the same sentence. When this has been done for the complete text, each individual sentence is then checked for the "score" of relatively high frequency words occurring in it, and sentences with the highest scores are then automatically selected, in textually-occurring order, and are printed out as an abstract, more properly an extract, of the document.

The automatic encoding of documents may be achieved either by taking the high ranking words of the selected sentences or by selecting the highest ranking of the words in the entire document as index entries. Luhn typically justifies these procedures as follows:

> "Of various automatic procedures for deriving typical patterns for characterizing documents, the systems here proposed are based on operations involving statistical properties of words ... It is held that the more often a certain word appears in a document the more it becomes representative of the subject matter treated by the author. In grading words in accordance with the frequency of usage within a document, a pattern is derived which is typical of that document and unique amongst all similarly derived patterns of a collection of documents. It is proposed that the more similar two such patterns are the more similar is the intellectual contents of the documents they represent...
>
> "... The creation of an encoding pattern may consist of listing an appropriate portion of the words ranking highest on the word frequency list derived from a document. Experiments conducted so far on documents ranging in size from 500 to 5000 words have indicated that word patterns consisting of from ten to twenty-four of the highest ranking words furnish adequate discrimination and resolution for retrieval, sixteen such words being a likely average." 1/

At Wright-Patterson Air Force Base an automated information selection and retrieval system has been developed jointly by Air Force and IBM personnel (Gallagher and Toomey, 1963 [205]). It involves both auto-indexing and auto-abstracting techniques following the Luhn word-frequency-counting techniques. Pre-editing is applied to demarcate fields (e.g., title, author) and to flag certain text words, particularly proper names, for special treatment. Special treatment, over and above the frequency-based selection score, is also given to words in the title field.

On the abstracting side, modifications to the original Luhn formula involve segmenting sentences in terms of strings of both high and low valued words separated by either periods or continuous strings of low valued words, on the assumption that long consecutive strings of low value words should weight negatively. The automatic extract consists of the highest ranking 20 percent of the sentences subject to the restriction that no less than 7 and no more than 20 sentences should be selected. On the indexing side, the investigators report:

1/ Luhn, 1959 [371], p. 47.

"As it is currently run, the auto-indexing program selects about one word in ten as a keyword in articles of three thousand words or less. In articles longer than three thousand words it tends to pick about one word in fifteen. This high incidence of keywords naturally increases the amount of noise results returned by the query program, although good search strategy cuts them down considerably." 1/

As of October 1963, the system was reported to be fully operative although not as yet extensively tested in actual use. Gallagher and Toomey give illustrative auto-extract results on two tested papers, one being Luhn's own "Automatic Creation of Literature Abstracts". They give comparative results for manual versus machine selection of keywords as index or search terms with 88.6 percent agreement, the human indexers having selected, in 6 tests reported, 132 words and the machine method 117. Modifications under consideration include pre-edit flagging of terms in author and cited-reference fields for special weighting, setting the length of the abstract as a function of the total number of words in an item, and, in the search program, generating additional search terms by means of association factor techniques such as those suggested by Stiles.

To the basic approach of straight-forward word frequency counting, Luhn himself has suggested that improvements might be obtained from considering closely adjacent words, 2/ word pairs, 3/ and reference to vocabularies specific to a given field. 4/ Other possibilities are capitalized words and lookup against an inclusion list. He also suggests:

"If certain words could be given in their relationships to other words, more specific meanings may be identified by such combinations. These relationships may range from the mere co-occurrence of certain words within a phrase or sentence to the combinations of specific parts of speech." 5/

Various investigators have proceeded to explore these and other possible improvements, including incorporation of relative frequency information, use of information about distances between high-ranked significant words, word pairs and word n-tuples,

---

1/ Gallagher and Toomey, 1963 [205], p. 51.

2/ Luhn, 1959 [384], p. 10.

3/ Luhn, 1962 [373], p. 11.

4/ Luhn, 1959 [384], pp. 8 and 10.

5/ Ibid, p. 5.

and other devices to improve detection of significant clues to subject content. Representative examples of such work will be discussed below. In addition, investigators abroad have developed modifications to the basic Luhn word frequency approach which appear to be necessary when it is applied to languages other than English. 1/

Thus, for example, Purto reports various investigations conducted by V. A. Argayev and V. V. Borodin and by himself with respect to Russian language documents. 2/ Purto notes first that the Luhn method as applied to Russian language materials selects sentences which, while having the largest "significance coefficients", were not those most essential to the meaning and further that: "an abstract in Russian made by Luhn's method results in a choice of sentences not conveying basic information and not logically connected with each other." 3/ The reasons for such failure he attributes to the fact that words with different frequencies are considered equally important within a sentence for sentence selection purposes and to the lack of consideration for semantic and grammatical connectivity between significant words and between sentences. He then discusses several methods for determining connectivity, such as the rule that the sentences most closely connected with each other will be those in which the greatest number of the same significant words occur. 4/

A somewhat different example of difficulties occurring when the basic Luhn technique is applied to material in languages other than English is given by Levery. He describes a study of thirty French texts concerned with the development and manufacture of glass. He reports as follows:

> "While we followed the classical idea that a relationship between the frequency of a word and its significance exists, the fact that we worked with French texts forced us to discount the value of frequency alone.
>
> "French authors generally do not like to repeat the same words, and they vary their vocabulary... It was necessary to combine the frequencies of words with the same meanings or related to the same idea."
>
> "A dictionary of synonyms was constructed... (and) different versions of the same word had to be regrouped." 5/

1/ Note, however, that in the automatic abstracting program at Thompson Ramo-Wooldridge, small-scale experiments suggest that automatic abstracting is as feasible for other Indo-European languages as for English, (1963 [603], p. ii). Also, at the Centre d'Etudes Nucléaire Saclay, automatic extraction experiments are being applied to texts both in French and other languages, see National Science Foundation's CR&D report No. 6, [430], p. 20.

2/ Purto, 1962 [484]. He refers to a report "The problem of automatic abstracting and a means of solving it", by Argayev and Borodin, apparently available only as a typescript dated 1959.

3/ Ibid, p. 3.

4/ Ibid, pp. 3-4.

5/ Levery, 1963 [359], p. 235.

3.3.2 Frequencies of Word n-tuples - Oswald and Others

The first alternative to the basic Luhn word frequency approach in automatic abstracting techniques to be actively explored was apparently that of Oswald and his associates. (Oswald et al, 1959 [459]; Edmundson et al, 1959 [180]). Like Baxendale, Oswald was interested in word pairs and word groups, particularly compound-noun and adjective-noun compositions, as more revelatory of meaning than single words. Unlike Baxendale, however, he was interested in the word group itself as selection criterion, whereas she had used word group or phrase clues for the selection of (usually) single indexing terms. Differences between their two approaches, both representing very early efforts in the field, are summarized by Edmondson and Wyllys as follows:

> "Oswald's experiment in automatic abstracting differs from Luhn's and Baxendale's techniques in that it combines the notion of significance as a function of word frequency and the notion of significance as a function of word groupings, by employing juxtapositions of significant words as the basic unit for measuring the importance of a sentence...
>
> "It may further be observed that Baxendale's exhibited indexes are made up of single words rather than word groups, in spite of the strong case she makes for using groups...
>
> "Baxendale's work is concerned solely with the automatic construction of indexes; she does not extend her treatment of word significance into the area of automatic abstracting." 1/

Oswald's "multiterms", however, were intended to overcome, in the areas of both automatic indexing and automatic abstracting, at least some of the difficulty that concepts are often expressed in compound nouns, word pairs, and longer groups of words consisting of n-tuples of substantive words or of phrases. The result of considering both word frequency and word-group frequency is that in Oswald's selection-groups it is usually the case that only one word of the group has an individually high frequency but the co-occurrence feature heightens the significance of the relatively lower frequency words with which it appears. Thus, for automatic indexing, Oswald proposed significant word groups as indexing terms, and his criteria for selection of sentences to be included in machine-generated extracts are similarly based on the number of significant groups in the sentences chosen.

Other investigators who have stressed the importance of word pairs and longer groups as necessary to reflect concepts include Bar-Hillel (1959 [33]), Black(1963 [64]), Clark (1960 [123]), Doyle (1959 [165]), and Salton (1963 [519]). Doyle says succinctly that "when a phrase, or some other aggregation of words, stands for a single idea, its frequency in a document ought to interest us more than the frequencies of its component words." 2/ Salton considers it desirable to use word groups rather than individual words

---

1/ Edmundson and Wyllys, 1961 [181], pp.231-232.

2/ Doyle, 1959 [165], p. 11.

for purposes of identifying document contents and to use data on the joint occurrence of words in the same sentence or similar contexts as grouping criteria. Clark points out in particular that the use of ordered pairs and longer sequences of words to express a single concept may be highly characteristic of the special technical language used in a specific subject field, and notably those of the social sciences. 1/

Others who have explored word n-tuples as selection criteria for automatic extraction operations include such investigators as Szemere, Levery, and Yakushin. Szemere reports an investigation of 39 Swedish patent specifications in the field of switching circuits looking for significant word-pairs, with emphasis on noun-adjective combinations (1962 [591]). The objectives of a project headed by Levery at IBM - France have been reported as follows:

> "A series of experiments is planned in the fields of automatic indexing of technical texts and technical vocabulary analysis.
>
> "A statistical method will be tested to determine the degree of closeness in meaning of words. The method will consist of studying the pairs of words which appear together in the majority of texts and calculating a coefficient of correlation from the frequencies. Such work will result in a standard list of notions frequencies for a particular kind of information.
>
> "Starting from this list, new experiments will be made so as to obtain a list of keywords representing each text. The method will use statistical comparison between the distribution of frequencies of notions contained in a text and the standard distributions obtained for the entire corpus." 2/

Yakushin(1963 [654]) develops a variation of the word-pair principle in which he looks for those pairs where the words are, or suggest, names of objects, such as "table-leg". He suggests, further, that so-called "basis nouns" can be established for a given scientific field and entered into an inclusion dictionary, which also contains codes for the lexical classes to which the word can belong and codes for determining whether or not the word can join with another as a "basis term". Machine routines are then suggested to develop whether or not given terms are jointly part of the same text, whether one textually precedes another in a given text, whether or not there is a "nomenclator" pair. Depending upon the frequency of occurrence of identical or semantically related nomenclator constructions, it is claimed that subject concepts can be detected. That is:

> "The method is founded on the finding in a text of so-called basis terms, established by list, and of the words which explain them. These explanatory words, which in different contexts refer to one basis term, are grouped and ordered according to definite rules into a subject concept." 3/

1/ Clark, 1960 [123], p. 460.

2/ National Science Foundation's CR&D report no. 11, [430], p. 118.

3/ Yakushin, 1963 [654], p. 16.

3. 3. 3 Relative Frequency Techniques - Edmundson and Wyllys, and Others

The first comprehensive critique of word frequency approaches to automatic extracting and indexing was undoubtedly that of Bar-Hillel (1959 [33], 1960 [34]), followed closely by Edmundson and Wyllys (1961 [181]), who themselves have experimented with various alternative or improved methods for obtaining measures of word significance by statistical analysis. These critics have been in agreement both on many points of specific criticism and on suggested possibilities for amelioration of observed difficulties, especially in terms of considering relative word frequencies within a particular subject field. In addition, several other investigators independently proposed a relative frequency approach at about the same time. 1/

Some typical expressions of opinion on the importance of relative frequency criteria are as follows:

> "Let me propose here a system of auto-indexing which, to my knowledge, has never been publicly proposed before in this form and which seems to me superior to any other system I have heard of ... Assume that ... we are given a list of the average relative frequencies of all English 'words' ... It would then be possible, for any given document, to rank-order all the 'words' occurring in this document according to the excess of their relative frequency within the document over their average relative frequency. By some mechanically implementable standard or other, an initial segment of this list is selected as the index-set." 2/

> "Very general considerations from information theory suggest that a word's information should vary inversely with its frequency rather than directly, its lower probability evidencing greater selectivity or deliberation in its use. It is the rare, special, or technical word that will indicate most strongly the subject of an author's discussion. Here, however, it is clear that by 'rare' we must mean rare in general usage, not rare within the document itself. In fact it would seem natural to regard the contrast between the word's relative frequency f within the document and its relative frequency r in general use ... as a more revealing indication of the word's value in indicating the subject, matter of a document." 3/

1/ Compare, for example, Kochen, 1963 [327], p. 7: "The idea of contrasting words which occur frequently in a document against the frequency of this word in the background language for purposes of selecting index terms seem to have been suggested first by Bohnert and the author, then described in more detail by Edmundson and Wyllys, and tested empirically by Damerau. Something similar was suggested even earlier by Bar-Hillel." See Bar-Hillel, 1962 [35], p. 418, footnote, with respect to himself, Edmundson, and Bohnert. See also, however, Doyle 1962 [163], p. 388: "Edmundson and Wyllys were probably the first to publicly advocate contrasting word frequencies within a document to word frequencies within a given field and using these relative frequencies as criteria for scoring and selecting sentences."

2/ Bar-Hillel, 1959 [33], pp 4-8-9.

3/ Edmundson and Wyllys, 1961 [181], p. 227.

"We naturally find that the words of greatest interest are those for which there exists the greatest contrast between general usage frequency and local (within the article) usage frequency." 1/

"Luhn has bypassed syntactical analysis by taking advantage of the information content of the most frequently used topical words in articles ... Edmundson et al take a further step in a desirable direction by bringing in information from outside the article being analyzed: words and terms are given greater topical value as the contrast increases between the frequency of use within the article and the rarity of general usage." 2/

"A further refinement of the process of automatic analysis would be the development of special sets of reference frequencies for special fields of interest. This would have two benefits: it would become possible to classify documents as to field, and it would become possible to note the significance of words which are frequent in the document and frequent in a very large reference class c_o of literature (i.e., these words would not be significant with respect to c_o) but which are rare in the special field. For example, the word 'emotion' might be too common in general usage to seem significant, but frequent occurrence of the word would stand out in a paper on electronic circuitry (e.g., of a robot) when compared with its frequency in general electrical engineering literature." 3/

"One of the ... goals is to investigate a relative-frequency approach to the categorization of documents... For this investigation it will be necessary to develop sets of reference frequencies for words used in different subject fields. It was suggested by Edmundson and Wyllys that these sets of reference frequencies, when developed, could be used to categorize a document as belonging to a particular subject-field, by means of measuring the degree of matching (e.g., with the chi-squared test) between the proportional frequencies of words in the documents and the sets of reference frequencies." 4/

Two points in the comments quoted above appear especially worthy of note. The first is that of introducing at least some measure of reference to material other than the individual author's own choice of linguistic expression and specific terms. We shall discuss this factor in more detail in a later section of this report. The second point, derived in part from the first, is the specific suggestion of movement away from purely derivative indexing by machine in the direction of automatic assignment indexing and automatic categorization or classification.

1/ Doyle, 1959 [165], p. 9.

2/ Doyle, 1961 [169], p. 3.

3/ Edmundson and Wyllys, 1961 [181], p. 228.

4/ Wyllys, 1963 [653], p. 10.

Actual experiments in application of relative frequency techniques to automatic extracting processes have been pursued since 1959 by various investigators. Edmundson and Wyllys and Damerau (1963 [148]) were certainly among the first. Edmundson and Bohnert were engaged in experimental investigations at Planning Research Corporation in 1959, [1] and the following year Edmundson, Oswald, and Wyllys worked on the auto-indexing and auto-extracting of the 40,000 words of text contained in nine articles in the subject field of missilery. [2] Wyllys has continued work on relative frequencies (1963 [653]). At the System Development Corporation Doyle, in some of his work, has also explored the relative frequency approach (1961 [161]). An example in Europe is work reported by Meyer-Uhlenried and Lustig, where significant keywords from abstracts are used not only as indexing terms directly, but by means of keyword lists and micro-thesauri can also be used to assign documents to specific subject fields (1963 [417]).

3.3.4 Significant Word Distances

Another technique that has been investigated for the improvement of automatic extraction operations based on the statistics of word frequencies is that of distances between significant words. The desirability of attaching greater weight to n-tuples of immediately adjacent words and to the co-occurrences of words within the same sentence has been mentioned previously. Savage, in relatively early work developing some of the initial proposals of Luhn, considered intra-sentence distances between significant words as follows:

> "... The criterion is the relationship of the high-frequency words to each other, rather than their distribution over the whole sentence. Consequently, it seems reasonable to consider only those portions of sentences which are bracketed by high-frequency words and to set a limit for the distance at which any two such words shall be considered as being significantly related ... An analysis of many sentences and many documents indicates that a useful limit is four or five non-significant words between any two high-frequency words." [3]

Doyle has also noted the tendency of words that are in fact highly related in a content-revealing sense to co-occur in the same sentence or as quite direct neighbors. The same investigator has also suggested that word distances can be used to provide "clustering" effects that might, for example, sort out the possibly different topics covered in introductory or background discussions, the main text, and various appendices. [4]

1/ National Science Foundation's CR&D Report No. 5, [430], p33; Bar-Hillel 1962 [35], p. 418.

2/ National Science Foundation's CR&D Report No. 6 [430], pp 43-44.

3/ Savage 1958 [521], p. 4. Later related work has included a method for generating auto-extracts which adds to the high-frequency word sentence scores a correction factor for the number of words in gaps between such words. (See Rath et al, 1961 [493].)

4/ Doyle 1961 [166], p. 7.

Related research efforts in more general areas of linguistic data processing suggest inter-sentence distances as criteria for the selection of words and word groups in automatic indexing and abstracting processes. In natural language text searching, for example, the work of both Swanson (1960 [587], 1961 [586], 1963 [583]), and of Maron and Ray [1] suggests that limitation of searching to a four-sentence span would eliminate a number of irrelevant responses to search requests specifying the joint occurrence of two or more words.

Swanson's findings indicated that if two words or phrases contained in the search request were found in textual proximity within these limits, they were highly likely to bear a semantic relationship that is what was intended by the requester. Applying the four-sentence proximity criterion, it was found that the amount of irrelevant material retrieved by the text searching system could be reduced by 60 percent without serious loss of relevant information. [2] Black cites the four-sentence proximity criterion and notes further that it might be used also to retrieve only a paragraph or similar small portion of the full text, reducing the amount of material to be read by the user, perhaps by as much as 90 percent. [3]

Artandi, in her book-indexing studies, suggested as a topic for further investigation the possibility that proximity of index term candidates as derived from the same section of the text could serve to improve the quality of the indexing. Since her computer program checks for duplicate potential entries occurring on the same page, this feature could be used for further analysis, on the assumption that the number of occurrences of the same entry for the same page is an indication of the importance of the discussion of the subject on that page. [4]

3.3.5 Uses of Special Clues for Selection

Intra- and inter-sentence distances between words are relatively crude examples of clues to selection of words and word-pairs which, because of their implied relationships, may be especially significant for indexing, sentence extraction, or document categorization. They can be quite readily detected by machine, but the implication that physical proximity is a good measure of significant co-occurrence is often false. Other clues which can be detected equally well, mechanically, are those which have to do with position and format.

1/ Ray, 1961 [494], p. 92.

2/ Swanson, 1963 [583], p. 9, 1961 [586], pp. 298-299.

3/ See Black, 1963 [64], p. 20 and footnote: "The figure 90 percent is derived from experience in previous experiments, wherein the amount of relevant material was scanned and a subjective judgment was formed that the relevant material was actually about 10 percent of the total verbiage retrieved. That is, about 10 percent of each document contained the relevant material; 90 percent of the document was of no relevance but the document as a whole was relevant."

4/ Artandi, 1963 [20], p. 47.

Such obvious positional clues as occurrences of words in titles, chapter or section headings, figure captions, have already been mentioned. To these can be added first and last sentences of paragraphs, 1/ or of first and last paragraphs as such. 2/ Wyllys observes that other criteria which are detectable in the text by straightforward machine procedures can be based on such features as italicization, capitalization, or punctuation. He notes, however, that such "editorial" criteria vary from journal to journal so that their usefulness would need to be related to the particular practices of individual journals. 3/

Somewhat more difficult for machine implementation, but certainly feasible in the present state of the programming art, is the use of specific semantic or syntactic clues. Here again, Luhn, Baxendale, and Edmundson and Wyllys all anticipate their critics and later investigators. Luhn recognized the fact that in at least some applications the characterization of documents by isolated words alone would fail to provide an effective degree of discrimination. He, therefore, suggested operations to establish word relationships, whether based on co-occurrences or combinations of specific parts of speech. 4/ Baxendale clearly uses both syntactic and semantic clues, detectable by built-in table lookups.

Representative suggestions by Edmundson or Wyllys or both as co-authors include the following:

> " ... We have in mind a glossary or dictionary of perhaps one to two thousand words that act either as cue words which signal the importance of a sentence or as stigma words that signal the insignificance of a sentence for purposes of abstracting." 5/

1/ See, for example, Wyllys, 1963 [653], p. 27: "One of the first published studies in automatic document-content analysis, that of Miss Phyllis Baxendale, brought out the importance of the first and last sentences in a paragraph as bearers of a good deal of the content of the paragraph." See also Marthaler, 1863 [399], p. 25.

2/ Compare Swanson, 1963 [580], p. 1: "...Some evidence exists to show that for short homogeneous articles title and first paragraph are nearly as good as full text."

3/ Wyllys, 1963 [653], p. 28.

4/ Luhn, 1959 [384], p. 5.

5/ Edmundson, 1962 [178], p. 11.

"The criteria for attributing significance to words ... may be positional (in virtue of their occurrence in titles or section headings), or semantic (in virtue of their relation to words like 'summary'), or perhaps even pragmatic (in the case of names of specialists mentioned in text footnotes, or bibliography ...

"A cataloguer or abstract-writer would naturally give more weight to a technical word that appears in a title, in a first paragraph, or in a summary. A machine can be programmed to do the same. It can be instructed to recognize the title by position and capitalization ... It can place first-paragraph indications... It can test every heading or subtitle for the words 'summary' or 'conclusions' and place a summary indication after each word in the summary paragraphs." 1/

"The statistical criteria ... by no means exhaust the potential clues to the representativeness of sentences. Among other plausible clues are certain words and phrases ... authors use words such as 'conclusion', 'demonstrate', 'disclose', 'prove', 'show', and 'summary' (and related forms of these) with high frequency in sentences that contain concise statements about the topic or topics of the article... The occurrence in a sentence of such a phrase as 'it was found that...', 'the experiment proves...', or 'the central problem is ...' would indicate probably even more sharply than any single word could that the sentence was likely to be highly representative of the topics..." 2/

3.3.6 Recent Examples of Mixed Systems Experimentation

It is quite obvious from the above samples of suggestions for the use of various special clues for automatic extraction, that improved systems will largely depend upon a mixture of means for determining subject-representativeness of words, phrases, and sentences. Many of the clues suggested by Edmundson and Wyllys are continuing to be explored, as mixed systems, at RAND 3/ and the System Development Corporation, (1962 [590]), for example. Two specific recent examples of mixed systems experimentation are the automatic abstracting experiment programs at Thompson Ramo-Wooldridge and the work involving detection of first incidences of nouns at the Harvard Computation Laboratory.

The TRW programs to investigate possibilities of computer generation of document auto-abstracts, involving both English and Russian language texts are based upon a combination of four different methods to measure significance and determine representativeness. These four methods are briefly described as follows:

"... The Key method has its source of machine recognizable clues the specific characteristics of the body of the document and is based on a Key Glossary of content words taken from the body of the document.

1/ Edmundson and Wyllys, 1961 [181], pp. 227 and 229.

2/ Wyllys, 1963 [653], p. 25.

3/ See National Science Foundation's CR&D report No. 11, [430], pp. 314-315.

"... The Cue method has as its source of machine recognizable clues, the general characteristics of the corpus that are provided by the bodies of the documents and is based on a Cue Dictionary of function words apt to appear in the body of a document.

"... The Title method has as its source of machine recognizable clues, the specific characteristics of the skeleton of the document, i.e., title, headings, and format, and is based on a Title Glossary compromising those content words found in the title, subtitles, and headings, but excluding certain words of the Cue Dictionary.

"... The Location method has as its source of machine recognizable clues, the general characteristics of the corpus that are provided by the skeletons of the documents and uses a Heading Dictionary of certain function words that appear in the skeletons of documents." 1/

The Harvard work involving detection of the first incidences of nouns as sentence selection and indexing clues is part of a larger-scale program for mechanized information selection and retrieval under the general direction of Salton (1961 [512], 1962 [513], 1963 [514] and [515]). The specific mixed system involving frequency data, syntactic identification clues, and positional criteria is primarily the result of investigations by Lesk and Storm (1961 [577], 1962 [358]). Related work takes advantage of computer techniques for predictive syntactic analysis and automatic dictionary lookup also under development at the Harvard Computation Laboratory (Kuno and Oettinger, 1963 [339], [340], [341]).

The Lesk-Storm experiments have involved investigations where the hypothesis assumed is that the points in a text where the author has first introduced a specific noun or nominal phrase, or where he has used, with higher frequencies, a combination of first-referred-to-nouns, are most likely to be especially indicative sections of text with respect to subject-content representativeness. The assumption is further, that areas in which specific "new" ideas, not mentioned previously in the text, are first introduced is particularly rich in topical-content concentration. 2/

The mixed-system emphasis followed by Lesk and Storm, however, is revealed in the following comments:

"It is not, of course, apparent that a count of initial occurrences of nouns ... is by itself sufficient to reveal areas of significant information content for purposes of abstracting or indexing. Accordingly, the method suggested here must be used together with other available means, and is not expected to provide by itself an acceptable abstracting algorithm." 3/

In their actual investigations, Lesk and Storm first made manual counts of initial noun occurrences in various sample texts, noting paragraph, sentence, and first incidence-of-word identifications. The computer was then used to carry out three distinctive tasks: (1) calculation of the number of new nouns for each sentence in the text;

1/ Thompson Ramo Wooldridge, 1963 [603], p. 1.

2/ Lesk and Storm, 1962 [358], p. I-6.

3/ Storm, 1961 [577], pp. I-1 and I-2.

(2) computation of functions proportional to the number of initially occurring nouns for each sentence, and (3) the preparation of a normalized graph for initial noun occurrences by plotting the functional values against each sentence in the text. 1/ Sentence selection can then proceed by processes to detect "peaks" on the graph, using a relative criterion or weighting function to minimize the effect of high first-noun counts in the beginning sentences of a paper.

Trials were made with a number of different weighting formulas, and the best of these involved the obtaining of moving averages of first-noun counts over several adjacent sentences. A particular formula covering a span of seven sentences gave results that appear to emphasize contextual effects and to reduce the effects of a particular single sentence with a large number of new nouns, such as a listing of proper names. The resulting abstracts are quite lengthy (e.g., comprising 20 percent or more of the original text), and contain some relatively uninformative sentences. The investigators think that the results with respect to satisfactory abstracting are inconclusive but provocative. They also conclude that the possibilities for indexing are more immediately promising: "Most key definitions are retained in the successful summaries, and the vocabulary reflects the topics covered in the texts." 2/

Other examples of mixed-system experimentation, especially involving the use of syntactic and semantic considerations, include the work at the General Electric Computer Department under Spangler, and work by Jacobson and Plath. In the Phoenix laboratories of General Electric, a KWIC type indexing program can be applied both to titles and to running text and a contemplated extension is intended to "generate indexes by means of word analysis, taking into consideration syntactic and semantic aspects of text lines". 3/ Jacobson describes rules for machine determinations of same-meaning occurrences of words which may be homographic and for selection of descriptors for indexing simple paragraphs by choosing words occurring at least twice with a high probability of having the same meaning. 4/ Plath reports:

> "Although sentences occur in which the key term or phrase lies buried deep down in the structure, preliminary observations indicate that there are many others in which the semantic hierarchy closely parallels that of the syntactic structure. This suggests that more sensitive vocabulary statistics for purposes of automatic abstracting may be obtainable by considering only words occurring in positions above a predermined cut-off level in the sentence structure. Alternatively, one might count occurrences of words on each level, and then multiply by a fixed weighting factor in each instance before taking the overall totals." 5/

1/ Lesk and Storm, 1962 [358], pp. I-2, I-4 ff.

2/ Ibid, p. I-31.

3/ National Science Foundation's CR&D Report No. 11, [430], p. 21.

4/ Jacobson, 1963 [292], p. 191-192.

5/ Plath, 1962 [474], p. 190.

3.4 Quality of Modified Derivative Indexing by Machine

Most of the modified derivative indexing techniques that have been proposed to date have few or no indexing results to provide comparative data for purposes of evaluation. Moreover, those techniques which are primarily directed to the generation of document abstracts rather than indexing terms have been reported to date with a paucity of actual examples. 1/ One of the main reasons for this lack of product-effectiveness data is unquestionably the high cost and difficulty of obtaining substantial corpora of representative document text in machine-readable form. For the most part, the few examples of automatic abstracts produced by machine are sadly lacking in pertinency, relevancy, 2/ and in continuity for scanning or reading by comparison with conventional human abstracts, whether prepared by author, editor, volunteer specialist in the subject field, or professional documentalist.

A few studies have been made for a somewhat larger numbers of examples of "auto-abstracts" with respect to differences between several different machine-extraction formulas, random sentence selections, and sentences extracted manually. A project conducted by IBM's Advanced Systems Development Division for the ACSI-matic program, (1960 [289], 1961 [290]), involved 70 to 90 articles on military intelligence items. The comparisons were of "auto-abstracts" as against titles, full texts, "pseudo-auto-abstracts" comprised of the first and last 5 percent of the sentences of each text, and sets of sentences selected randomly, without reference to conventional types of manually prepared abstracts and without respect to the quality as such. Similarly, Thompson Ramo Wooldridge data (1963 [601]) on machine-extracted and randomly-extracted, sentence sets compare these "abstracts" against manual selection of 25 percent of the sentences of each item, rather than against a conventional type of abstract.

There are however, almost no data available on the possible results of using sentence and word-group extracting techniques, applied to machine-usable texts, to the development of indexing entries rather than to the generation of substitutes for document abstracts. For this reason, as well as because discussion of the difficulties of evaluation in general will be deferred to a later section of this report, the question of the quality of modified derivate indexing will be briefly considered below, largely in terms of non-quantitative judgments.

First and foremost, as has been noted previously, is the objection that word-indexing typically produces redundancy, scatter of references among synonyms and near-synonyms, inclusion of many irrelevant entries at high page and user-scanning costs, omission of

---

1/ Purto expresses regret that the studies of Agrayev and Borodin, intercomparing results of human abstracting, use of Luhn's method, and their own modification, used only a single paper (1962 [484]). Storm, (1961 [577]), evaluating the initial noun occurrence technique as a measure of sentence and index-term extraction significance, reports results for only two papers, both by Quine. Only nine articles, with no more than 40,000 words of text in toto, were used by Edmundson, Oswald and Wyllys in their 1960 experiments ([180]).

2/ Compare, for example Lesk and Storm, 1961 [358], pp. I-29 and I-30 as follows: "A final problem is the ambiguity that may arise by removing two sentences from context; two sentences alone do not always permit comprehension. Worse yet, the meaning may actually be inverted upon removal from context. For example... a quote is selected which an unsuspecting reader might think the author supports, when he is really attacking the position."

many properly indexable topics or points of interest because the authors did not emphasize them or used new and unusual terminology to describe them, failures to achieve consistency both of reference and index-vocabulary control for the papers of more than one author, and the like.

Additional difficulties are engendered, for word indexing by machine from text as against word indexing by people, because of complexities required in programming to achieve recognition of even such simple indicia as endings of sentences, 1/ inconsistencies of capitalization, 2/ and misspellings. 3/ Context distinctions between multiple meanings of homographic words are even more difficult. Difficulties in achieving good indexing quality are increased if only titles are used; those of keystroking and machine cost requirements increase as the amount of input material grows.

For these reasons, early criticisms such as those of Bar-Hillel are largely as pertinent today as they were when statistical techniques for computer generation of document extracts and index terms were first proposed. For example:

> "There can be no doubt but that computers are in a position to select out of the words or word-strings occurring in the encoded form of the original document those words or strings which fulfill certain formal, statistical conditions, such as occurring more than five times, occurring with a relative frequency at least double the relative frequency in general... However, it is ... unlikely that the set obtained thereby will be of a quality commensurate with that obtained by a competent indexer. First, there will be serious difficulties as to what is to be regarded as instances of the same word ... Second, there arises ... the problem of synonyms. Third, and most important, this procedure will yield at its best a set of words and word strings exclusively taken from the document itself." 4/

On the other hand, there are many situations where, because of time factors or lack of conventional indexing resources, even unmodified derivative indexing by machine is itself of value and therefore modifications to improve the quality of results, whether made by man or by machine, may be well worthwhile. As Anzlowar claims: "The increasingly widespread KWIC indexes ... can save so much in time and effort that they surely deserve better than the somewhat haphazard 'slash-dash-ing' now done in most in most instances as the only cerebral operations thereon." 5/

1/ See Luhn, 1959 [384], p. 22: "Amongst the difficulties encountered in the processing of machine readable texts, inconsistencies in the use of punctuation marks, compounds, capitals, spacing and indentations have been a problem way out of proportion with respect to the simple functions these devices stand for. For instance, even with the aid of a dozen different tests performed by the machine, the true end of a sentence cannot be determined with certainty."

2/ See Artandi, 1963 [20], pp. 52ff, on problems of capitalization of proper names.

3/ See Wyllys, 1963 [653], p. 15.

4/ Bar-Hillel, 1962 [35], pp.417-418.

5/ Anzlowar, 1963 [16], p. 104.

Modifications to derivative indexing techniques that tend toward normalizations of terminology and word usage, and increasingly sophisticated proposals for machine use of syntactic, semantic, and contextual clues hold out the promise of transition to more truly "subject" indexing and to automatic assignment indexing systems.

4. AUTOMATIC ASSIGNMENT INDEXING TECHNIQUES

Answers to the question of whether indexing by machine is possible are actually dependent in part on how the question of whether what can be achieved by machine is or is not properly termed "indexing" is answered. If "indexing" is defined as being more than the mere extraction of words from titles, abstracts, or text, then automatic derivative indexing, even when augmented by various modifications, normalizations, and editings, does not provide affirmative evidence. In the case of concept-oriented definitions of indexing, the question becomes one of whether or not automatic assignment indexing is possible. Experimental evidence suggesting that it is will be presented in this section.

We should note first, however, that just as there are differences of opinion as to what "indexing" means so there are similar differences, with respect to whether or not it represents concepts rather than extracted words. There are also a number of conflicting definitions of what is meant by "indexing" in contradistinction to "classifying". For some, the latter difference is related to questions of the number of labels or surrogates assigned to a single item to represent its subject contents, ranging from the assignment of a single subject category in a classification scheme involving mutually exclusive classes to the assignment of a number of terms or descriptor each standing for one of a number of aspects of the subject. For our purposes, however, we shall regard both the case of indexing with a number of descriptors and that of classifying to a single category or subject heading as being within the province of automatic assignment indexing, reserving the term "automatic classification" for the case where the machine is used to establish the classification or categorization scheme itself.

Actual experiments in automatic assignment indexing by Borko, Borko and Bernick, Maron, Salton, Stevens and Urban, Swanson, and Williams will be discussed briefly below. These discussions are generally in chronological order with respect to first reporting of results, except that the Salton-Lesk-Storm work reflects a somewhat different principle of assignment from the methods using clue word approaches and it is therefore described after these others have been discussed. Some of the similarities and differences between the various methods are then indicated. A brief final subsection covers related assignment indexing proposals for which experimental data is not available or has not as yet been reported in the literature.

4.1 Swanson and Later Work at Thompson Ramo-Wooldridge

Research on fully automatic indexing as well as on full text searching and retrieval at the Ramo-Wooldridge Corporation has been reported as being under way at least as early as the spring of 1958. 1/ As described elsewhere in this report, experiments in search and retrieval based upon full natural language text had used as test items short articles in the field of nuclear physics. In additional experiments representing a preliminary "clue word" approach to possibilities for automatic indexing procedures, some of this same material was used.

1/ National Science Foundation's CR&D rept. no. 2, [430], p. 32.