

COMPUTER GENERATED INDEX TO THE URANTIA BOOK

CONSIDERATIONS:

Permission to undertake, authorization by Foundation, cooperation and coordination with Foundation.

Selection of committee to oversee preparation of index.

Decisions regarding finances, retail cost, distribution, announcements of availability.

Method of financing, whether there shall be any return of profit to private investors if that is the method used.

Decisions regarding criteria of index:

Length of book, probably same size as URANTIA Book.

Method of selecting words to be indexed.

Different ways ~~xxx~~ to index.

Ways to index: classes of "indexability":

- 1. Number of occurrences (a number)
- 2. Page references (1257b)
- 3. KWIK: (Key word in a line of context.)

Proposed classes: (These seven are the only reasonable combinations.)

- A. Number only
- B. Number and all Page
- C. Number and all KWIK
- D. Number and some Page
- E. Number and some KWIK and remaining page
- F. Number and some KWIK and some page
- G. Number and some KWIK and no page

1978

Assume 20,000 word vocabulary in U.Book

Then just to list the stats for each word (on a separate line)
would take 20,000/50 lines per page = 400 pages

For some words, which occur in phrases, they need not be listed
in both areas, but whole sections could be deleted from one
listing and a reference included to cross-reference the other.

Example: Under listings for "thought", the reader would be referred
to "adjuster" if he is interested in Thought Adjuster.
the Cross-reference would be in its proper alphabetical
location:
ountain top and thought about how he would live out th
thought adjuster: see "adjuster"
re more than he thought possible, and his young mind t

If a word appears more than once on a page it could be listed 1234(5)
91(2)
2097(3)

1978

Thorndyke Counts (Based on random samples of books, magazines, Encyclopediae, etc.)

# words Exact	# words Approx.	per word Occurrences per million	total occurrences per million
5209	5000	1	5000
2503	2500	2	5000
1442	1500	3	4500
1064	1000	4	4000
890	1000	5	5000
	1500	6-7	10000
	1000	8-9	8500
	1000	10-13	11000
	1000	14-18	16000
	1000	19-29	25000
	1000	30-49	40000
952	1000	50-99	70000
1069	1000	over 100	806000
	19,500		1,000000

Selected frequencies
per million

the -----50,000
of, a, and, to -----25,000
it -----10,000
course----- 500
problem----- 250

O/M	#W (Thorndyke)	#W (Urantia Book) estimated		
1	5000	100	Being a unified work, it is probable that the Urantia Book has a smaller vocabulary than that of a random count. Despite the numerous unique, new words in the Urantia Book, there are no extraneous words, technical terms, etc. Those words which are used are therefore used more frequently. Hence this adjustment of the Thorndyke count.	
2	2500	1000		
3	1500	1000		
4	1000	1000		
5	1000	1000		
6-7	1500	1500		
8-9	1000	1000		
10-13	1000	1000		
14-19	1000	1000		
19-29	1000	1000		
30-49	1000	2000		
50-99	1000	2000		
100+	<u>1000</u>	<u>2000</u>		
	19,500	15,600		

1979

ACTUALLY GENERATING THE INDEX

The computer would do most of the menial work, but human decisions would be required at various points along the line:

1. ~~In~~ setting up the overall approach (This the the preliminary step, and this paper represents the ~~first effort.~~)
2. ~~Refining the overall approach.~~ (This would be done by the selected committee, using this paper as a take-off basis.)
3. Arranging with the computer company and working with them to obtain the several products involved at the various stages of the ~~production~~ process.
4. The first step with the computer is to have the URANTIA ~~Book~~ Book "read" by an optical scanner, a mechanical character recognizer which encodes the book onto computer tapes or discs.
5. The next step is to instruct the computer to read it's tape and produce a WFL or word frequency list. This lists every ~~to~~ word occurring in The URANTIA Book in alphabetical order along with a number which represents the number of times the word occurs in the Book.
6. Now comes the most tedious and important step in the human decision process, the assignment of words to "index classes" The index classes have been established in steps one and two.
 - a. The first part of this step is a gross division ~~into~~ into classes. Some words can be assigned easily to their final classes. Others may require committee decision. Others will require interim KWIK printouts from which to make further decisions. Namely classes D through G.
7. The words for which interim KWIK printouts are needed are sent back to the computer for KWIK printouts.
8. The interim words are further perused and finalized, assigned to D,E,F,G as appropriate, with appropriate instructions.
9. Send all instructions to computer and have it print the finished index.
10. Bind and reproduce and ~~disseminate~~ distribute the indexes.

1979

CLASSIFICATION CATEGORIES (Manual)

By Importance: By Frequency:

a = essential	1 = 1
b = important	2 = 2-8
c = indexable	3 = 9-39
d = incidental	4 = 40-100
	5 = 100+

PRINTOUT MODES (Automated)

= List the word and the number of times it occurs only.
 R = Word + # + all REFERENCES by page and line. E.g.: 1773m
 K = Word + # + all KWIK (Key word in context) with refs.
 KR = Word + # + specified KWIK's (with page and line references) +
 remaining occurrences referenced by page and line.

Note: KWIK, or key-word-in-context, means that the word is
 printed as it appears in the book, with as much context
 as will fit on one line.

CLASSIFICATION MATRIX WITH SUGGESTED PRINTOUT MODES

IMPORTANCE	F R E Q U E N C Y				
	1	2-8	9-39	40-100	100+
Essential	K	K	K	K	KR
Important	K	K	K	KR	KR
Indexable	K	R	R	R	
Incidental	K	R	#	#	#

Instructions to person assigning CLASSIFICATION CATEGORIES

Assign K to all single frequency words
and All essential words with Frequencies under 100
and All important words with frequencies under 40

Assign KR to all essential words with frequencies over 100
and all important words with frequencies over 40

Assign R to all indexable words with frequencies from 2 to 100
and all incidental words with frequencies from 2 to 8

Assign # to all other words, that is,
all indexable words with frequencies over 100
and all incidental words with frequencies over 8

RESULTS OF DELPHI TECHNIQUE USED AT WORKSHOP ON COMPUTER POTENTIALS
 LED BY CHICK MONTGOMERY AND DAN MASSEY (TOTAL 30 PARTICIPANTS)
 AT SUMMER STUDY SESSIONS AT LAKE FOREST, ILLINOIS, AUGUST, 1980

The following lists the Consensus of Estimates by the afternoon participants, based on preliminary estimates of morning participants:

<u>"How many years until....?"</u>	<u>Estimated number of years</u>
First authorized translation published	3
Business and word processor (computer) at Chicago Headquarters . . .	3½
First machine-produced index to U.Book published	4½
Newsweek or Time reports on the Book or Movement	4½
Newsweek or Time reports on attack by organized religion on U.Book	6
Authorized translation in two languages	6½
U.Book announcements (ads) in bookstore trade magazines	7½
First machine-assisted translation of U.Book	10
U.Book available on pocket micro-cassette player	10½
U.Book available on pocket micro-fiche reader	10½
Major movie based on Part IV	11½
Pocket U.Book available on micro-thin paper	13
1,000,000 Books in print (10x)	15
3,000 study groups world wide (10x)	17½
130 Societies world-wide (10x)	20
100 employees at Chicago (10x)	20
Translations available in 10 languages	22½
U.Book advertized (announced in paid ads) in general media . . .	25
Translations available in 100 languages	37½
50% of world population use same currency	50
50% of world population read one language	62½
50% of world population are of same general race	1500

MACHINE TRANSLATION (MT) OR MACHINE ASSISTED TRANSLATION (MAT)

APPLICABILITY TO THE URANTIA BOOK

- * PRE-EDITING SIMPLIFIED (Because the subject document is a single work, of finite length, in the English language)
- * CONTINUOUS REFINEMENT POSSIBLE IN SUCCESSIVE ITERATIONS
- * RELATIVE ABSENCE OF SLANG, JARGON, FIGURES OF SPEECH, AND ITS UNIQUE LINGUISTIC PRECISION RENDER THE URANTIA BOOK PARTICULARLY SUSCEPTIBLE TO MACHINE TRANSLATION
- * SUPPRESSION OF HUMAN INTERVENTION
- * FIXED AND FINITE VOCABULARY (facilitates preparation of the program "dictionary" and allows for full disambiguation)
- * SINGLE SOURCE LANGUAGE AND MULTIPLE TARGET LANGUAGES
- * ENABLES THE FOUNDATION TO QUICKLY "OCCUPY THE FIELD" IN THE INITIAL TRANSLATION
- * RAPID ACCELERATION OF MACHINE TRANSLATION TECHNOLOGY AND PROGRAMMING IN NEXT FEW DECADES

ADVANTAGES OF MACHINE TRANSLATION AND MACHINE INDEXING

TECHNOLOGY EVOLVING RAPIDLY; ~~THE~~ ~~EVOLUTION~~
 AVAILABILITY: CRUDE PROGRAMS AVAILABLE NOW

ACCURACY: MACHINES DON'T ERR, NO TYPOS

CONSISTENCY: NO HUMAN JUDGEMENT BEYOND SETTING UP PROGRAM
SAME TREATMENT OF WORD OR ISSUE THROUGHOUT TEXT.

ECONOMY: PAGE-READER ENCODES BK. PERFECTLY IN A FEW HOURS
 TRANSLATION & INDEXING FUNCTIONS AUTOMATIC.
 PRINT OUT CAN BE CAMERA-READY.

EASY
 TRANSLUCIBILITY: (TRANSLATIONS, ~~NOT~~ ^{NOT} INDEXES) ~~ONCE~~ ~~PREPARED~~
 PRE-EDITING (INDICATING PARTS OF SPEECH, CLARIFYING FIGURES
 OF SPEECH) AND ENCODING SOURCE NEED ONLY BE DONE ONCE
 FOR ALL LANGUAGES.

FIDELITY: HUMAN INTERVENTION IS ^{KEPT TO A MINIMUM,} ~~LIMITED~~ ~~AND~~ ~~POSSIBILITY~~ ~~FOR~~
 COLORING OR MISINTERPRETING OR DOWNSTEPPING MEANINGS
 IS THEREFORE ~~LIMITED~~ ~~REDUCED~~ ~~SUBSTANTIALLY~~.

FEASIBILITY OF SCOPE: ONLY ONE TEXT IS INVOLVED: THE JOB IS HANDLEABLE.
~~FEASIBILITY~~ THIS MAKES PRE-EDITING AND CONTINUAL ~~BE~~ LATER REFINEMENTS
 OF PRE-EDITING ~~POSSIBLE~~ ~~FEASIBLE~~.

OCCUPY THE FIELD: ESTABLISHED COPYRIGHT IN TARGET LANGUAGE(S).
~~PRECLUDE~~ ~~OR~~ ~~OF~~
 CURBATE ~~OR~~ INCENTIVE FOR UNAUTHORIZED TRANSLATIONS

PROVIDE BASIS FOR PRODUCING LATER ~~EDIT~~ IMPROVED EDITIONS

ADVANTAGES UNIQUE TO URANTIA BOOK MACHINE TRANSLATION

1. Only one text, finite job. Text has long term viability, no revision possible for centuries, therefore can continually polish translations with no fear of obsolescence.
2. Text is of a precisely worded nature, nearly devoid of figures of speech, and therefore easy of translation.
3. Text may be (we don't know yet) especially designed to facilitate translation into other tongues. There are signs that this may be so. Example: frequent use of long strings of prepositional phrases beginning with the word "of".

THE PROCESS OF GENERATEING A COMPUTER INDES

PROGRAMMING THE COMPUTER TO GENERATE THE INDEX

~~C~~ ~~U~~
V

Encode the book: keypunch operation produces machine-readable tape.

Produce Word Frequency List (WFL): use existing software.

Check WFL for accuracy (look at single occurrence words)

Correct machine-readable tape.

Produce corrected WFL:

Classify all words according to

a. ignore: classification clear

b. produce complete KWIK listing for analysis

Receive trial KWIK list for b. words.

classify all words a and b according to the following categories:

~~xx~~ EL (essential: low frequency)

EH (essential: high frequency)

N (non-essential)

CU Classification unclear (N vs. E)

Request KWIK for EH and CU words

Classify CU words

Edit all EH ~~KWIK~~ KWIK lists.

Program for final printout. (Stat, all cite, all KWIK, specKwik+rem.
cite.)

Receive final printout index.

Programming for final printout:

12 - ...
13 - ...

CRITERIA FOR COMPUTER-GENERATED INDEX

- should
1. It ~~shall~~ be the same size as the Urantia Book.
 2. It should be generated from a machine readable tape,
So that further applications of the tape can be made.
 3. It should contain all criteria for its generation
clearly stated in a preface.
 4. It should be as useful as possible, given the limitations.
 5. It should not attempt to duplicate the manually prepared
index.
 - ~~6. The master tape should be thoroughly accurate.~~
 7. It should be produced at minimal cost to the Foundation.
 8. It should be reproducible at the same cost as the
Urantia Book.
 9. There should be a smaller edition published which contains
only the essential and important words, without any
context entries, which could be reproduced for a few
dollars or less.
 10. All KWIK lists generated during the developmental phases
of the index should be retained by the Foundation and
made available to persons doing detailed, specialized
research.
 11. Each human decision phase of development should be *Committee.*
performed in its entirety by a single ~~individual whose~~
~~interests, accomplishments, professions, biases, etc.~~
~~shall~~ ~~be~~ ~~briefly~~ ~~described~~ ~~as~~ ~~well~~ ~~as~~ ~~possible.~~
Human decision phases are: classification by importance, and
selection of specified KWIK's. Assignment of frequency and
mode determination are not human decision phases.
 12. If possible, it should be handled by a firm interested in
donating its services to a worthwhile non-profit organization.

1982

ENCODING THE URANTIA BOOK IN MACHINE-READABLE FORM

POTENTIAL APPLICATIONS

- * RE-SETTING TYPE FOR FUTURE PRINTINGS
 - In same face and format
 - In new type-face and/or new format (e.g., large type)
- * PRODUCTION OF INDEX AND/OR CONCORDANCE
- * PRODUCTION OF MACHINE TRANSLATIONS (MT) OR MACHINE-ASSISTED TRANSLATIONS (MAT)
- * PRODUCTION OF URANTIA PAPERS IN OTHER THAN PRINTED BOOK FORM
 - Braille
 - Voice-readers for the blind or visually impaired
 - Micro-fiche pocket video readers
 - Micro-cassette voice readers
 - Future technology
- * LINGUISTIC ANALYSIS

1982

PRODUCTION OF INDEX AND/OR CONCORDANCE USING AUTOMATED WORD PROCESSING

APPLICABILITY TO THE URANTIA BOOK

- * TECHNOLOGY IS HERE NOW AND PROVEN. A VARIETY OF PROGRAMS AND FORMATS FOR BOTH INDEXES AND CONCORDANCES AVAILABLE NOW.
- * COST IS GREATLY REDUCED COMPARED TO MANUALLY PREPARED INDEXES.
- * SUPPRESSION OF HUMAN INTERVENTION IS ALMOST TOTAL.
- * CAN BE USED TO ASSIST IN PREPARING A MANUAL INDEX IF DESIRED.
- * ACCURACY AND COMPLETENESS ARE ASSURED.
- * WOULD ALLOW THE FOUNDATION TO IMMEDIATELY "OCCUPY THE FIELD" OF THE ENGLISH LANGUAGE INDEX WITH A SUPERIOR PRODUCT IN TERMS OF COMPLETENESS, FIDELITY AND USABILITY.

1982

THE USES OF A COMPUTER-GENERATED INDEX TO THE URANTIA BOOK

1. To aid in the preparation of manually-produced indices.

The manually-produced index currently in the final stages of preparation could be checked against the computer-generated index to assure inclusion of all relevant references. ~~And, it may be expected that as world conditions change, revised editions of the Urantia Book index will be desirable, in the preparation of which a computer-generated index would be an invaluable aid.~~

2. To provide a more comprehensive index for serious students.

Dedicated scholars, researchers and investigators into the Urantia Book will not be entirely satisfied with a manually-generated index as their only reference aid. Notwithstanding the limitations of a computer-generated index, it does have the following advantages:

- a. Completeness. Each and every reference to those words selected for inclusion in the index would be listed.
- b. Infallibility. It is possible to assure with absolute certainty that no unintended omissions occur.
- c. Objectivity. The human judgements involved are necessarily limited and can be precisely stated in the preface to the index.