A TRANSPORTATION INFORMATION HANDLING SYSTEM
(UN SYSTEME D'INFORMATION APPLIQUE AUX TRANSPORTS)

Pierre A. Yansouni*
Canadian Transport Commission
Research Branch
Computer Systems and Research Support
Ottawa, Ontario

* Acknowledgements: The author has merely undertaken to report the results of a team effort by the staff of the Computer Systems and Research Support Directorate of the Canadian Transport Commission; particularly those of A. Bailey, J.-A. Johnson, B. Litvack, L. O'Connell, A. White.

ABSTRACT

The Research Organization in the Canadian Transport Commission has developed a Transportation Information Handling System and implemented a prototype version to a point where it is ready to become operational. The system falls into the input coordinated category, with a hierarchial classification structure reflecting a generalized model of the transportation process and its interactions with the economic, social and environmental systems. Classification and retrieval are accomplished by relating information and data to one or several paths in the classification structure. A "hybrid" design, however, allows the user to revert to a "natural language" (keyword) search at any level of the classification structure, restricting the search to a subset of records which size is determined by the level of the structure to which the path(s) has been extended. The computerized retrieval package allows unrestricted boolean association of paths and keywords.

USER REQUIREMENTS AND POTENTIAL BENEFITS

Information required to conduct the research function in the Canadian Transport Commission (CTC) is transferred in three major forms: bibliographical, numerical data, and verbal communication. A large segment of the bibliographical information and the data actually used is unpublished and often of a confidential nature. Verbal communication is conditional to the identification of centers of expertise (individuals, research groups, consultants, etc.).

A survey of the research staff showed a widespread conviction that experienced personnel have no difficulties monitoring and accessing relevant published information, using conventional library catalogues, indexes, abstracting services, etc. Therefore, benefits of a system referencing this type of information reside mainly in shortening the time required to gather material, and in facilitating the transition from one area of expertise to another.

A much stronger requirement existed for referencing unpublished material; particularly, a large amount of factual information acquired by the CTC staff in conducting transportation studies, and numerous data bases established within the CTC and other government and private organizations with related activities. Experience showed that this very valuable (and costly) information rested very often in the hands of small groups or even individuals who had little time to document and publicize it. In addition to making valuable information accessible, a system referencing the content of divisional and private files, coupled eventually with a centralized repository of duplicate information, would provide an

internal communication mechanism, and lessen the dependence of accumulated expertise on the mobility of the personnel. Centers of expertise such as research centers, consultants, etc. were to be treated in much the same way as unpublished information, and referenced in the system.

The idea of referencing material accumulated by the users in relation to their functions was taken a step further by seeking cooperation of the users in referencing this material themselves, in order to retain and communicate expert qualification of the information as to its quality and possible functional applications. It is essentially this consideration which led to the design of an input coordinated system based on a detailed hierarchial classification structure described in the following section. Potential benefits of this system could only be completely attained if its maintenance (input) eventually became in large part a cooperative effort of its users.

CLASSIFICATION STRUCTURE AND CONCEPTS

All CTC activities can be related to a general model of the transportation process reflecting its interactions with the economic, sociological and environmental systems, as well as the interrelations between transportation modes. Each of these elements is in itself a subsystem characterized by attributes (variables and functional relationships) describing each subsystem and its effects on the others. The transportation expert is normally familiar with all these attributes and can easily relate a piece of information to a subset of attributes describing the transportation system. These attributes, aggregated in categories to a desired level of definition, and properly referenced by means of a controlled vocabulary and an associative grammar (the classification structure described below), provide the mechanism for classifying and indexing all relevant elements of information.

Each attribute relates to three major concerns: the geographical location, the modal aspect, and the topical content. These will be called for convenience the geographical, modal and topical dimensions of the attribute. Each dimension can be further categorized into subsets, i.e. the modal dimension is categorized into the AIR, RAIL, ROAD, etc. modes (see Figure 1). Similarly the topical dimension is categorized into SUPPLY, DEMAND, PERFORMANCE, and the geographical dimension is categorized into WORLDWIDE, MULTINATIONAL, NATIONS, etc. This process of subdivision can be continued for each subset to the desired level of definition, i.e. AIR is divided into VEHICLES and FACILITIES and so on. It finally yields a hierarchial classification structure of which segments are illustrated in Appendix A. Figure 1 represents the major subsets of each dimension: the details of selected subsets being represented in Appendix A.

Note that on Figure 1, a fourth dimension is shown which differs from the previous three in that it is used to qualify the form and value of the information rather than its conceptual content. For the purpose of referencing information this dimension is used in the same way as the three others, however, for simplicity it is ignored in the following explanations. In this context an attribute (or an aggregated category of attributes) is represented as a point in a three dimensional space and is defined by its components in the geographical, modal and topical dimensions. This space is called the classification space.

Any piece of information is referenced in the system by identifying the related set of attributes and representing each of these points in the space by its respective components in the classification diagrams. As an example consider referencing a document reporting on the volume of passenger travelling by air on scheduled airlines from Montreal, and the volume of cargo travelling by air on scheduled airlines out of Toronto. This

document generates two points in the space, each represented by its respective components (see diagrams in Appendix A).

- point 1
  G1: Urban; Urban Center; Montreal.
  M1: Air; Operations; Commercial; Scheduled; Passenger.
  T1: Performance traffic utilization; Operations; Industry; Distance volumes (passenger or vehicles); Passenger volume.

- point 2
  G2: Urban; Urban Center; Toronto.
  M2: Air; Operations; Commercial; Scheduled; Cargo.
  T2: Performance traffic utilization; Operations; Industry; Distance volumes (cargo).

Written definition of each component by a string of words can simply be replaced by a path drawn on the appropriate diagram. Appendix A assembles the diagrams representing point 1. Note that several components could be represented on each diagram with the restriction that all components in a set of diagrams must, when combined, generate valid attributes. As an example, assume that the following attributes are related to a document:

G1, M1, T1
G2, M1, T1
G3, M1, T1
G4, M2, T2
G4, M3, T2

Components G1, G2, G3, G4, M1, M2, M3, T1, T2 could generate 24 points in the space and to avoid creating nonexistent references the diagrams must be grouped in two sections

- section 1

  G1 M1 T1
  G2
  G3

- section 2

  G4 M2 T2
  M3

From each section valid attributes are generated by taking all possible combinations of components.

Details of the referencing and coding procedures are described in (1) and (2). Completion of the record requires the preparation of an abstract, and if there is more than one section as defined above, the preparation of special sectional abstracts relating to one or several of these sections. See Appendix B for a completed record.

In the geography diagram the "boxes" define a type of node. The next level of disaggregation will be the enumeration of all possible nodes, which obviously cannot be explicitly represented on the diagram. The same circumstances apply to "boxes" referring to commodities, companies and associations in the supply, demand, and performance diagrams. For the purpose of retaining the level of definition corresponding to specific nodes, companies, etc., extensive lists have been prepared, which must be used to specify on the diagrams the appropriate node, commodity, etc. necessary to completely define the component of the relevant attribute (i.e. Montreal, on the geography diagram of appendix A). These lists are open ended and could be extended as necessary.

Information related to one or several attributes is retrieved by specifying their components (paths) on the appropriate diagrams in the same way as when referencing information.

Two types of difficulties are inherently related to the design of the system: the number of records associated with one attribute could become too large, and documents could be misclassified, in which case they are "lost" in the system. To alleviate these difficulties the system allows the switch to a natural language (keyword) search of the set of records selected by one or several paths, at any level of the classification structure. The system searches the abstract, title, author, and date fields of the records. If a natural language search is opted for, without any structured search, the whole data base is accessed in this manner. i.e. If the components G1, M1, T1 under point 1 were used for a query, the subset of records retrieved could be searched more selectively by amending T1 to read: T1; Business; Trips. Addition of these keywords will retrieve the records dealing with business trips.

DESIGN AND IMPLEMENTATION OF THE COMPUTER RETRIEVAL SYSTEM

Each record is converted to a machine readable form, and is assigned a "document number" key to an index-sequential file in which the record is entered. This data base is not normally accessed on-line for reasons of economy. The retrieval system accesses, rather, a subsidiary file where only the document number, author, and title of each record are stored. These fields are displayed on-line as a result to each query. The user can eventually request off-line printout of selected complete records, or consult hard copies maintained in a card file with the same sequential key.

The design of the retrieval system is very efficient storage wise and is particularly suited to boolean combinations within a search. Consider first the mechanism allowing the structured search. The basic elements are an index-sequential file, and a set of bit stacks, one for each of the "boxes" in the classification structure, identified by the unique number at the bottom of each box (see Appendix A). For each record in the data base, consecutive segments are entered in the sequential file, corresponding one to one to the sections in each of the records. Each segment contains the respective document and section numbers (see Figure 2). Everyone of the bit stacks contains as many bits as there are segments in the sequential file. Conceptually, during initialization of the system, each segment retains temporarily the identification of all the boxes defining the components of the attributes in the corresponding section. A program reads the file, segment by segment and turns to one, in all the stacks identified by the box numbers in the segment, the bits placed in the same sequential order as the segment in the file. This design allows the retrieval program to identify each section and document numbers related to a given box, by simply counting the bits turned to one in the stack having the same box number. It can be shown that this design is more efficient storage wise than the inverted list approach, when the number of bits equalled to one in each stack is larger than approximately 300. Boolean association of boxes is equivalent to performing the same operation on the respective bit stacks. Groupings and associations can be specified using parenthesis. Definition of a path in the structure is obtained by a sequence of box numbers separated by logical AND. several paths can be defined simultaneously using logical OR and parenthesis. The operator AND NOT can also be used.

Extension of this design to include retrieval on keywords is straight forward. A stack of the same length is simply added for each keyword (see Figure 2). Since the keywords relate to a document rather than a section, all the bits corresponding to segments containing the same document number are either zero or one simultaneously. Keywords recognized by the system include all the lists of geographical nodes, commodities, companies and

associations, and all the terms in the classification structure. In addition, the system will recognize significant words from the searchable fields that are not in any of the previous lists. These are extracted automatically from the searchable fields by comparison to preestablished stop-word and keyword lists. The two lists are updated after examination of a residual list of words not recognized as stop-words or keywords by the system.

Interaction with the system is through three major commands. The QUERY command allows the user to enter a query as a string of keywrods and box numbers associated at will using logical operators and parenthesis. Each string is numbered automatically and this number can be used to introduce the string in another query. After eack query is entered, the number of sections retrieved is automatically displayed by the system. The DISPLAY command displays previous query strings one by one. The COUNTDOC command produces the number of records, rather than sections, that have been retrieved, and allows the display of their document numbers, titles, and authors, all together or through a browsing option. Error messages, promptings and a HELP command are designed to help inexperienced users. As an example, the query under point 1 above is displayed as it would be entered by a user: (using "," as the symbol for AND)

1/ 5,9,MONTREAL,25,26,33,42,56,906,907,912,917,938.
(5,9 = G1; MONTREAL = M1; 25,26,33,42,56,906,907,912,917,938 = T1)

An experienced user could use the sequence of boxes in the diagrams to abreviate his query when no ambiguity can be introduced, i.e.

1/ MONTREAL,25,42,56,907,912,938.

OPERATING STATISTICS

a- Data base

- Number of records, published	2,000
unpublished	2,000
- Number of sections	6,000
- Number of references	120,000
- Number of references per record	30
- Average number of characters per abstract	1,200
- Cost per record of referencing a document, using dedicated personnel	$10.50
- Cost per record for keypunch	$ 1.00
- Cost per record for processing	$ .50

b- Computer system

- Total on-line storage in bytes*	4,000,000
- Cost of CPU per hour of operation	$20.00
- Cost of remote connection per hour	$10.00

* The storage could be reduced by at least one third, using various measures such as implementing inverted lists instead of stacks for keywords related to less than 300 sections. However, at this preliminary stage it was judged unnecessary.

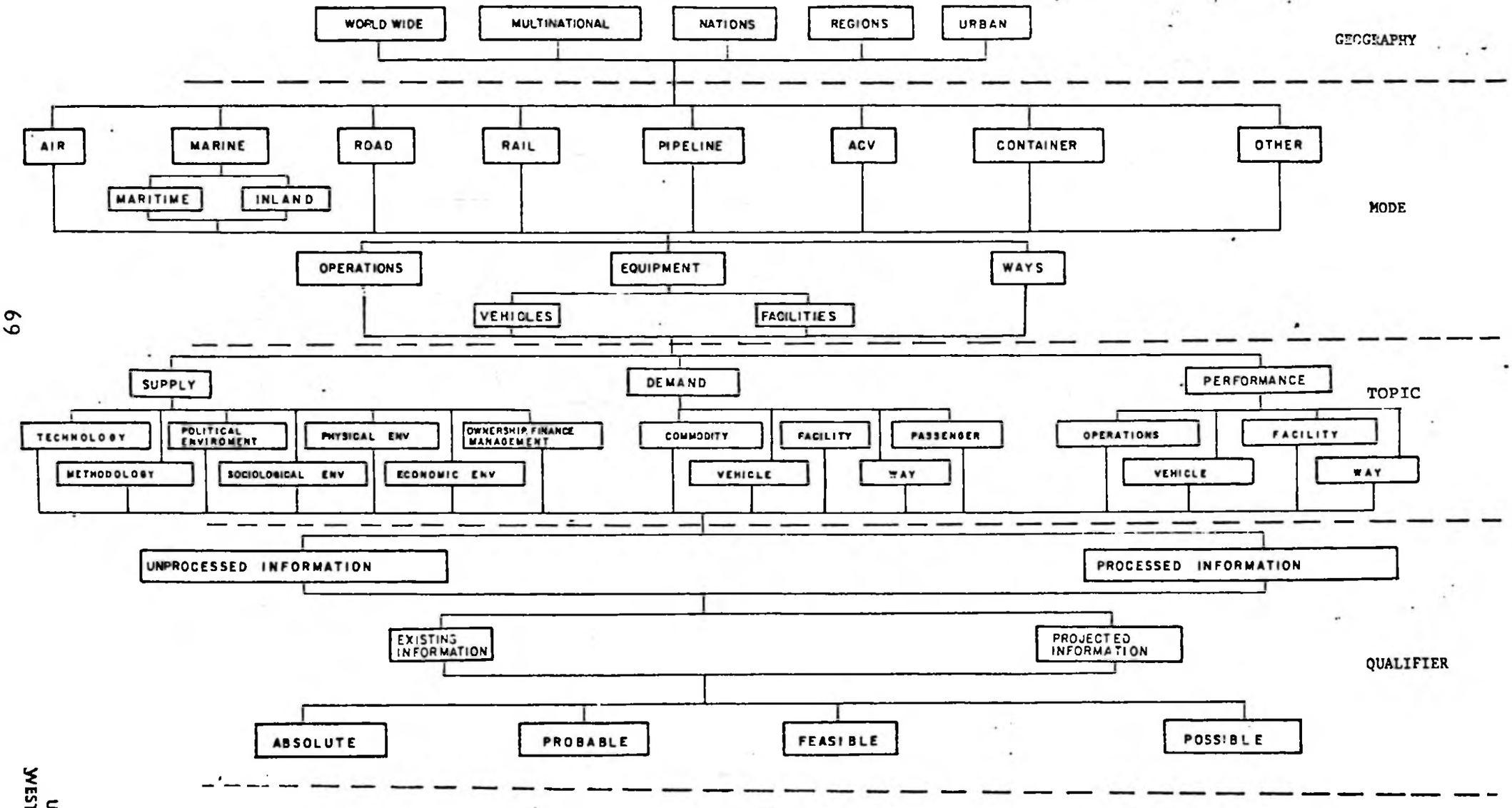
REFERENCES

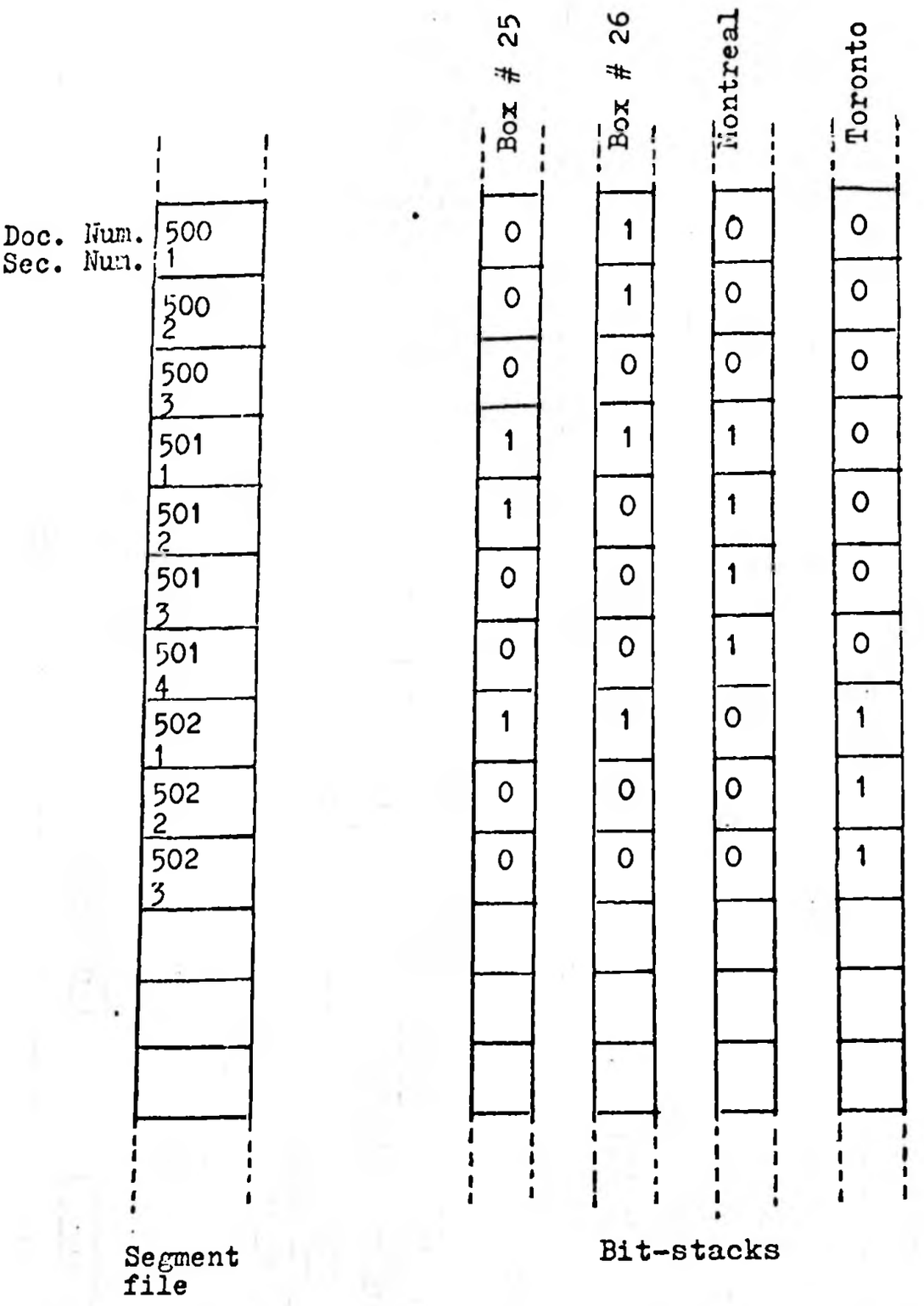
Reports

(1) CANADIAN TRANSPORT COMMISSION, Research Branch, 1975, Information Handling System, Report # 170, Volume 1.

(2) CANADIAN TRANSPORT COMMISSION, Research Branch, 1975, Information Handling System: A Manual Retrieval System, Report # 170, Volume 2.

Fig. 1

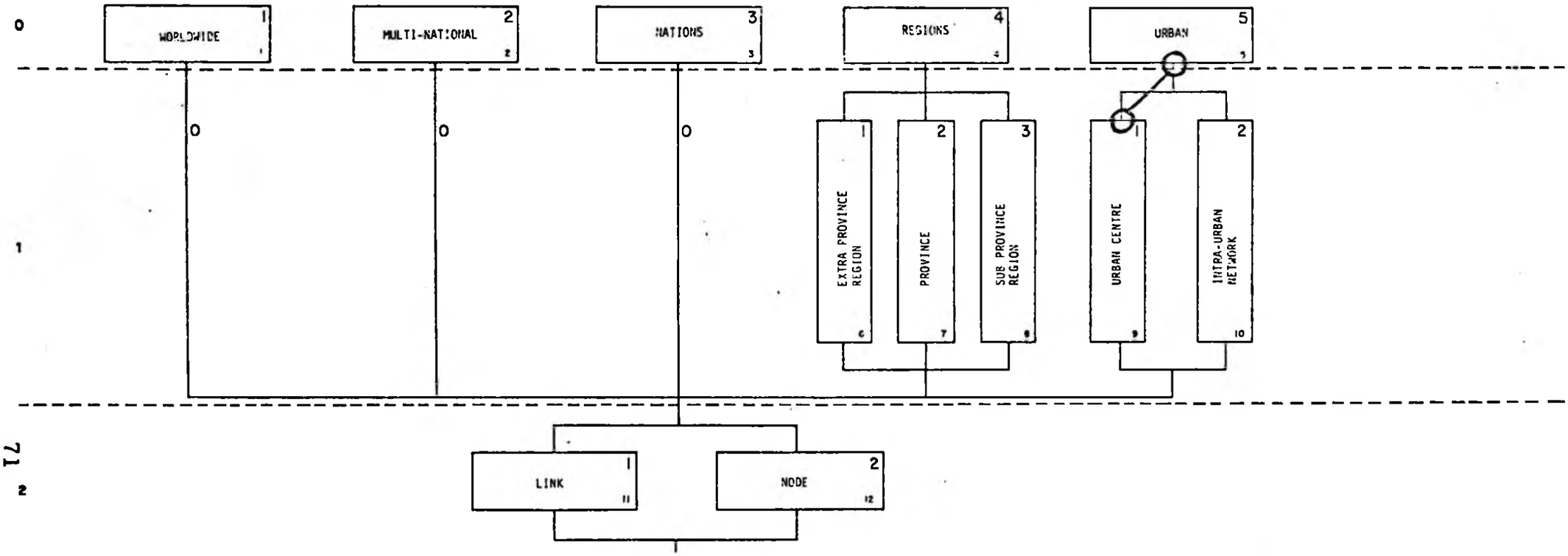




Note: Doc # 502 Sec 1 contains 25,26,Toronto.

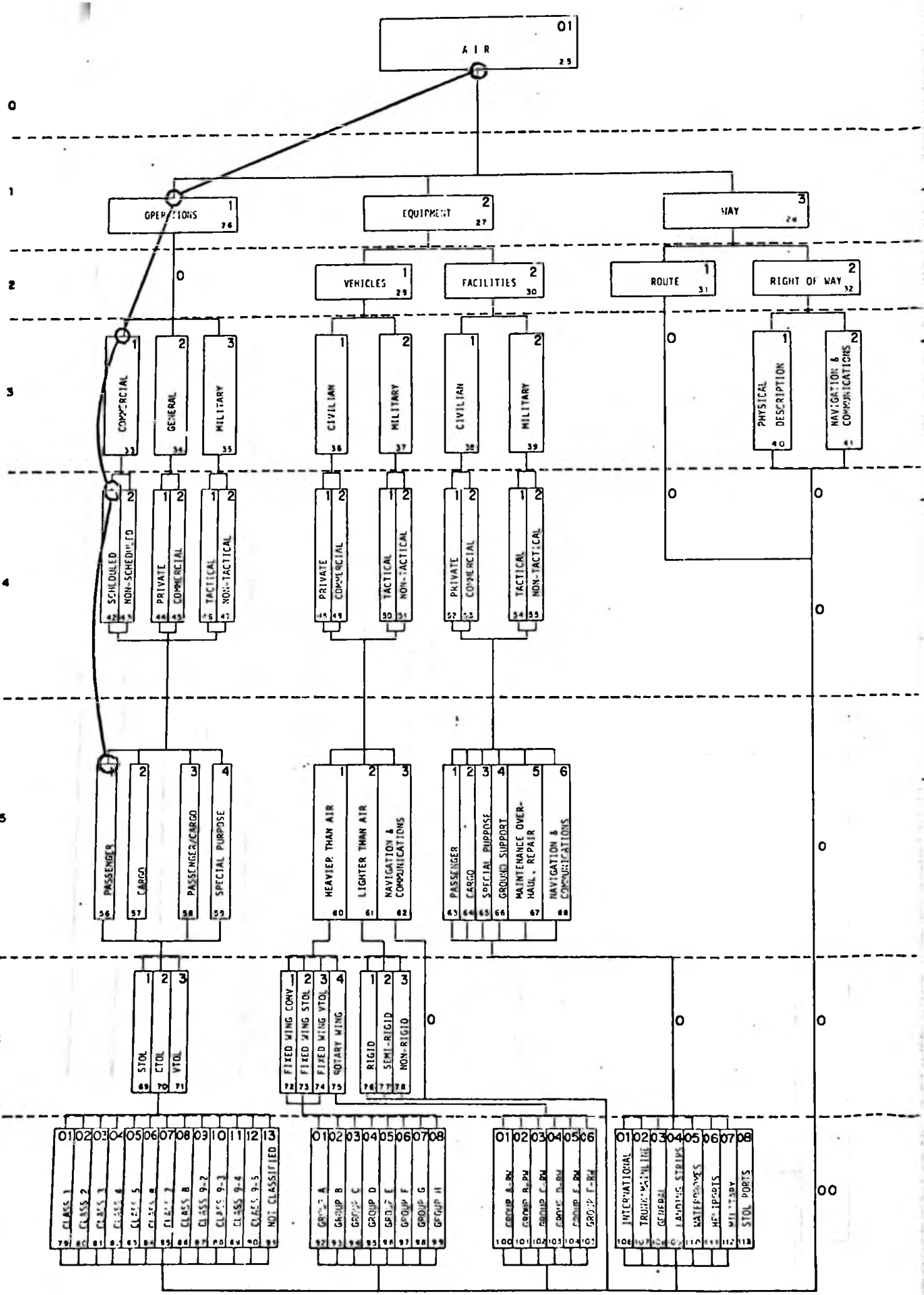
Fig. 2

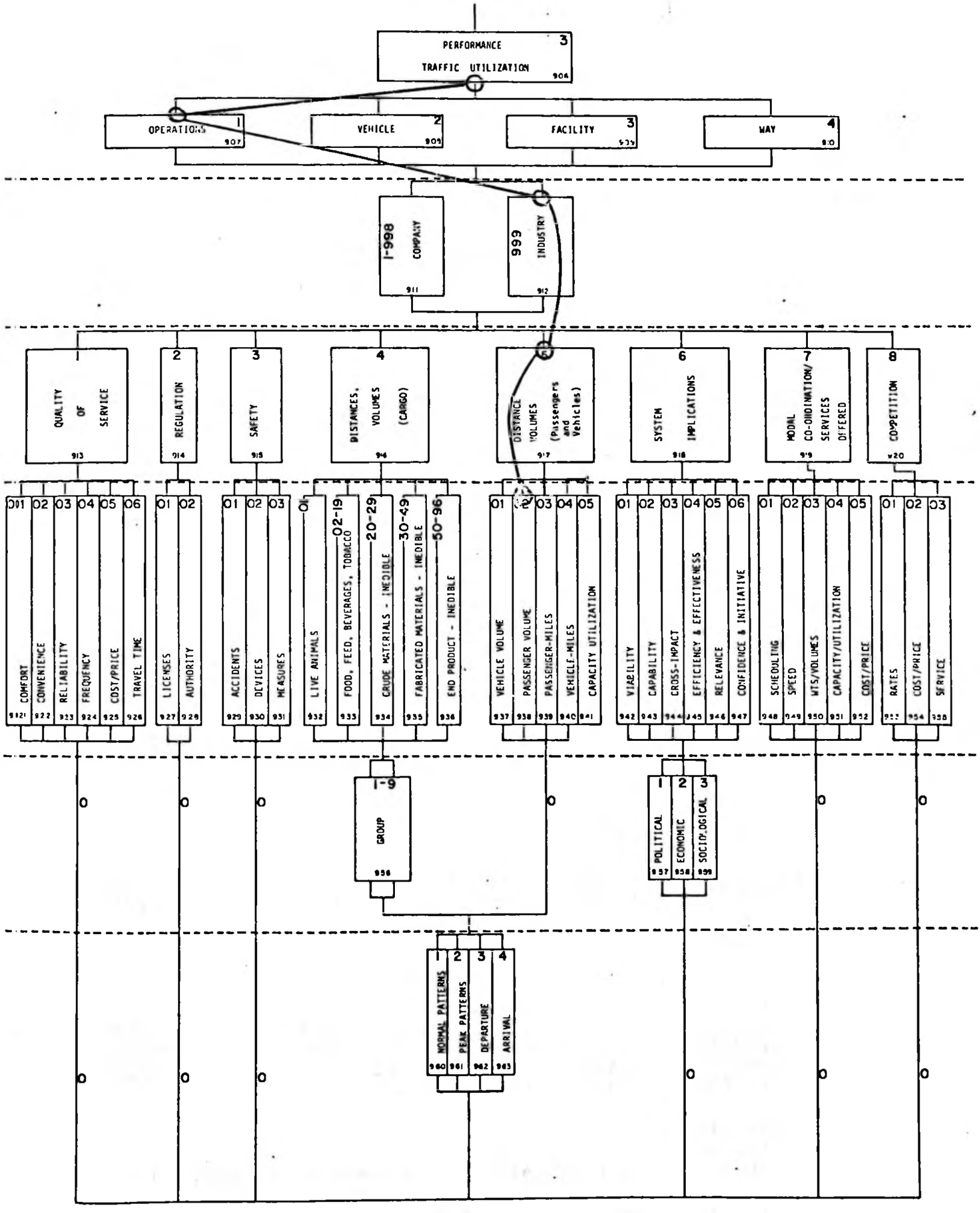
APPENDIX A



NODES OF INTEREST

Montreal


0
1
2
3
4
5
6
7
01
AIR
25
OPERATIONS
1
26
EQUIPMENT
2
27
WAY
3
28
VEHICLES
1
29
FACILITIES
2
30
ROUTE
1
31
RIGHT OF WAY
2
32
COMMERCIAL
1
33
GENERAL
2
34
MILITARY
3
35
CIVILIAN
1
36
MILITARY
2
37
CIVILIAN
1
38
MILITARY
2
39
PHYSICAL DESCRIPTION
1
40
NAVIGATION & COMMUNICATIONS
2
41
SCHEDULED
1
42
NON-SCHEDULED
2
43
PRIVATE
1
44
COMMERCIAL
2
45
TACTICAL
1
46
NON-TACTICAL
2
47
PRIVATE
1
48
COMMERCIAL
2
49
TACTICAL
1
50
NON-TACTICAL
2
51
PRIVATE
1
52
COMMERCIAL
2
53
TACTICAL
1
54
NON-TACTICAL
2
55
PASSENGER
1
56
CARGO
2
57
PASSENGER/CARGO
3
58
SPECIAL PURPOSE
4
59
HEAVIER THAN AIR
1
60
LIGHTER THAN AIR
2
61
NAVIGATION & COMMUNICATIONS
3
62
PASSENGER
1
63
CARGO
2
64
SPECIAL PURPOSE
3
65
GROUND SUPPORT
4
66
MAINTENANCE OVER-HAUL, REPAIR
5
67
NAVIGATION & COMMUNICATIONS
6
68
STOL
1
69
CTOL
2
70
VTOL
3
71
FIXED WING CONV
1
72
FIXED WING STOL
2
73
FIXED WING VTOL
3
74
ROTARY WING
4
75
RIGID
1
76
SEMI-RIGID
2
77
NON-RIGID
3
78
01 02 03 04 05 06 07 08 09 10 11 12 13
CLASS 1
CLASS 2
CLASS 3
CLASS 4
CLASS 5
CLASS 6
CLASS 7
CLASS 8
CLASS 9-2
CLASS 9-3
CLASS 9-4
CLASS 9-5
NOT CLASSIFIED
01 02 03 04 05 06 07 08
GROUP A
GROUP B
GROUP C
GROUP D
GROUP E
GROUP F
GROUP G
GROUP H
01 02 03 04 05 06
GROUP A-RW
GROUP B-RW
GROUP C-RW
GROUP D-RW
GROUP E-RW
GROUP F-RW
01 02 03 04 05 06 07 08
INTERNATIONAL
TRUNK/MAINLINE
FEEDER
LANDING STRIPS
WATERWAYS
HELIPORTS
MILITARY
STOL PORTS
00


PERFORMANCE
TRAFFIC UTILIZATION
3
904
OPERATIONS
1
907
VEHICLE
2
909
FACILITY
3
939
WAY
4
910
1-998
COMPANY
911
999
INDUSTRY
912
1
QUALITY OF SERVICE
913
2
REGULATION
914
3
SAFETY
915
4
DISTANCES, VOLUMES (CARGO)
916
5
DISTANCE VOLUMES (Passengers and Vehicles)
917
6
SYSTEM IMPLICATIONS
918
7
MODAL CO-ORDINATION/ SERVICES OFFERED
919
8
COMPETITION
920
COMFORT
CONVENIENCE
RELIABILITY
FREQUENCY
COST/PRICE
TRAVEL TIME
LICENSES
AUTHORITY
ACCIDENTS
DEVICES
MEASURES
LIVE ANIMALS
FOOD, FEED, BEVERAGES, TOBACCO
02-19
CRUDE MATERIALS - INEDIBLE
20-29
FABRICATED MATERIALS - INEDIBLE
30-49
END PRODUCT - INEDIBLE
50-96
VEHICLE VOLUME
PASSENGER VOLUME
PASSENGER-MILES
VEHICLE-MILES
CAPACITY UTILIZATION
VIABILITY
CAPABILITY
CROSS-IMPACT
EFFICIENCY & EFFECTIVENESS
RELEVANCE
CONFIDENCE & INITIATIVE
SCHEDULING
SPEED
MTS/VOLUMES
CAPACITY/UTILIZATION
COST/PRICE
RATES
COST/PRICE
SERVICE
1-9
GROUP
956
POLITICAL
ECONOMIC
SOCIOLOGICAL
NORMAL PATTERNS
PEAK PATTERNS
DEPARTURE
ARRIVAL

AUTHOR(S):

02142
DOC.NO.

TITLE: AIR CANADA'S TRANSCONTINENTAL & MIAMI MARKETS: A CLOSE-UP VIEW

PUBLICATION: AIR CANADA

DATE OF PUBLICATION: 01 01 74

SOURCE LOCATION: CTC-CI&P-L.M. O'CONNELL (BINDER-LOAD FACTOR#4)

AVAILABILITY: ON REQUEST

CONFIDENTIAL: NO X YES

SYSTEM ENTRY DATE: 25 08 75

FINAL CLASSIFICATION: NO YES X

OTHER INFORMATION:

DOCUMENT NUMBER: 02142

ABSTRACT: THIS REPORT EXAMINES THE IMPLICATIONS OF JUMBO JETS ON SOME OF AIR CANADA'S MAJOR ROUTES - MONTREAL-MIAMI & TORONTO-VANCOUVER. THE ANALYSIS CONSIDERS SPILL & LOAD FACTORS.

SECTION 1
Pp. 1-5
THIS SECTION DEALS WITH AIR CANADA'S TRANSCONTINENTAL ROUTE. COMPARISONS ARE MADE ON PROFITABILITY AND LOAD-SPILL FACTORS FOR 727, 747, AND L-1011 & SERVICE PROJECTIONS ON FUTURE TRAVEL ARE PRESENTED.

SECTION 2
Pp. 6-11
THIS SECTION DEALS WITH MONTREAL-MIAMI TRAVEL. AN ANALYSIS OF FIRST CLASS TRAVEL AND SPILL-LOAD FACTORS FOR AIR CANADA IS INCLUDED.