Hemalata IYER
School of Information Science and Policy
SUNY, Albany, NY, USA

Semantic Interpretation of Conjuncts: Boolean Transformations*

Iyer, H.: **Semantic interpretation of conjuncts: Boolean transformations.**
Int.Classif. 19(1992)No.2, p. 72-76, 5 refs.
This paper reports on an exploratory study of the semantic interpretation of conjuncts and their translation into Boolean search statements, using dictionary definitions. Rules were formulated based on syntactic and semantic analysis of the conjunctive phrases occurring in 160 natural language statements (NLS) of users information needs. This includes a set of transformational rules to accommodate variations in natural language expressions. A heuristic based algorithm, primarily intended to test the applicability of the rules on larger samples of NLS, was developed. Evaluation of the rules was performed by matching the output of the algorithm with the search formulation done by an expert online searcher. It resulted in an 81% match rate. (Author)

1. Introduction

The growing interest in end-user searching of online bibliographic databases has resulted in efforts towards designing front-end systems for translating natural language statements representing users' information needs to Boolean expressions. One of the problems in handling natural language is with conjunctions as they tend to introduce ambiguity. The conjunctions coordinate conjuncts that could include different kinds of grammatical constituents making their Boolean interpretation very difficult. Formulating a search expression for online searching in bibliographic databases essentially involves the process of identifying and combining the key concepts with the appropriate Boolean operators. The conjuncts in Natural Language Statements (henceforth referred to as NLS) of a user's information need very often represent the key concepts. Conjunctive phrases constitute important segments in NLS which can be used while formulating search expressions. The study explores the possibility of using conjunctive phrases for automatic formulation of search expressions from an NLS.

2. Objectives and Methodology of the Study

This is an exploratory study, undertaken to gain insights into the problems and issues involved in automatic Boolean interpretation of conjunctions 'and', 'or', 'but', occurring in NLS. Therefore it focuses on the analysis of conjunctive phrases occurring in an NLS; it explores the possibility of using dictionary definitions to determine the semantic similarity / dissimilarity of the conjuncts; as well as the subsequent translation of the conjunctions into the appropriate Boolean operators.

The central issue in this process is how to establish the semantic similarity between the conjuncts which could form the basis for the use of the appropriate Boolean operator, to combine the conjuncts in a search formulation. For this purpose the study uses dictionary definitions, adapting the techniques used by Chodorow [1] of automatically developing semantic hierarchies. Definitions are said to be constituted of 'genus' and 'differentia specifica'. Since the word that indicates the genus term refers to the class to which the defined word belongs, this represents the most essential property of the concept; whereas, the differentia specifica represent the properties of the concept that help to distinguish it from other concepts belonging to the same class. The genus part of the definition thus serves to assign the defined word to a class whereas the differentia specifica part of the definition helps to form the subsets within that class;

e.g. Copper: Copper is a soft reddish metal that is a simple substance, is easily shaped and allows heat and electricity to pass through it easily.

In this definition the word 'metal' is the genus term and the rest of the definition constitutes the differentia specifica.

Various attempts in the recent past have been made to use machine readable dictionaries in an information retrieval environment. Among the types of relations between terms / phrases, synonyms, relation and taxonomic relation are identifiable from the dictionary definitions [3,5]. Das-Gupta [2] in her exploratory study of Boolean interpretation of conjunctions strongly suggests the need for further study of this problem. The present research builds on her work and expands on the ideas given therein. However it is different in the analysis of the phrase patterns; the resulting rules; and it also extends over more types of phrase patterns than were derived from the analysis of the sample NLS.

A total of 268 NLS were collected from the State University of New York at Albany Library. These were search requests submitted by the users for online bibliographic database searching. They were screened to check for the presence of conjunctions. Out of these 268 NLS, 185 comprised conjunctions and 83 did not. A subset of 160 NLS from 185, was analyzed for this study. Table 1 gives a subject-wise breakdown of the NLS analyzed.

https://doi.org/10.5771/0943-7444-1992-2-72 - am 13.01.2026, 01:17:57. https://www.inlibra.com/de/agb - Open Access -

Humanities queries account for 5% and natural science queries to 3.75% of the total number ofNLS. The majority of the NLS were from the social sciences.

Definitions for conjuncts were taken from Longmans Dictionary of Contemporary English (LDOCE), and the genus terms were matched for their similarity. In case of a nonmatch, the definitions for the genus terms were taken and a hierarchy of the genus terms was developed. If the conjuncts wcre similar, the Boolean operator 'OR' was introduced and if they were dissimilar the Boolean operator 'AND' was introduced.

e.g. *Query:* "Aggressive behavior of handicapped children and adolescents"

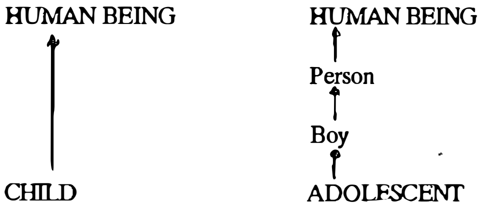
Definitions: In cases where the conjunct was a phrase, the definition for the headword of the phrase was taken.

Child, Children: A young human being of either sex, from before birth to the completion of physical development
Genus term = human being

Adolescent: A boy or a girl in the period between being a child and being an adult.
Genus terms = boy, girl

Boy: A young male person
Genus term = person

Person: A human being considered as having a character of his or her own, or as being different from all others.
Genus = human being



In this manner the definitions were traced via the hierarchy of genus terms in order to establish similarity / dissimilarity of the conjuncts. In this example, the conjuncts are similar; hence, the Boolean interpretation is as follows:

Aggressive behavior AND (Handicapped Children OR Adolescents)

The conjuncts appearing in the NLS represented several different grammatical constituents such as prepositional phrases, adjectival phrases, single words, pronouns etc.

Table 2 presents the distribution of the patterns of conjunctive phrases in the sample of 160 NLS.

Only 94 (58.75%) of the total NLS, given in Table 2 were categorized under various phrasal patterns and the corresponding phrasal rule applied. In this table, the order of the constituents of the phrasal patterns is not taken into consideration, for example, 23 occurrences in word-phrase pattern includes the phrase-word pattern also. The remaining NLS were covered by other rules such as, the comma rule, lexical rules etc. In this study 'word' refers to single terms such as alcohol, art, etc.; phrase refers to adjectival or adverbial phrases like 'Christmas parties'; and prepositional phrases consists of a preposition followed by a noun or a noun phrase. In many instances, it is taken to include the noun or the noun phrase preceding the preposition, such as 'special education of handicapped children'.

Semantic analysis of the conjuncts was done using the dictionary definitions from LDOCE. Rules were formulated for the Boolean interpretation, which were based on the analysis of the NLS. While developing the rules, it was found necessary to incorporate syntactic information for the Boolean interpretation. Essentially they involve semantic, syntactic analysis of the conjuncts and a set of transformational rules to accommodate the variations in the natural language expressions. An algorithm which is based on heuristics, was developed primarily for the purpose of extensively testing the applicability of the rules on NLS. The rules for handling different patterns of conjuncts follows.

3. Rules for the Algorithm

3.1 Comma Rule:

A, B, C, and D;

If commas are present and the last word is "anded" and the information needs statement ends there, then

-if C and D are similar then
A OR B OR C OR D

-if C and D are dissimilar then
(A OR B OR C) and D

-if commas are present and the information need statement continues beyond the last word.

e.g. A, B, and C in D.
"Art, beauty and aesthetics in literature", then
(A OR B OR C) AND D

3.2 Word-Word Rule:

If the conjuncts are single words this rule is applied. This requires semantic analysis of the conjuncts.

-if A, B are similar then
A OR B
e.g., Policy and programme
Policy OR Programme

-if A, B are dissimilar then
A AND B
e.g., Women and alcoholism
Women AND Alcoholism

Table 1. NLS Analysed: Subjectwise Breakdown

Soc. Welfare	Education	Public Adm	Crim. Justice	Lib Sc.	Humanities	Business	Poli. Sc.	Nat Sc.	Beh. Sc.	Others
35	41	11	6	10	8	3	4	6	20	16
21.8%	25.6&	6.8%	3.75%	6.25%	5%	1.87%	2.5%	3.75%	12.5%	10%

Table 2: Distribution of Conjunctive Phrasal Patterns

Phrasal Patterns	Number of Occurrences
Word-Phrase	23 (14.37%)
Phrase- Phrase	15 (9.37%)
Prepositional- Word Phrase	30 (10%)
Prepositional- Phrase Phrase	16 (3.75%)
Prepositional-Prepostional Phrase Phrase	6 (6.38%)
Word-Word	4 (2.5%)

Table 3 : Distribution of Matches

Phrasal Patterns	Matches	Non-Matches
Phrase-Word	11 (92%)	1 (8.33%)
Word-Phrase	10 (91%)	1 (9%)
Phrase-Phrase	14 (93%)	1 (6.67%)
Word-Word	4 (100%)	0
Prepositional-Word Phrase	16 (94%)	1 (5.88%)
Word-Prepositional Phrase	11 (85%)	2 (15.38%)
Prepositional-Phrase Phrase	4 (67%)	2 (33.33%)
Phrase-Prepositional Phrase	8 (80%)	2 (20%)
Prepositional-Prepositional Phrase Phrase	4 (67%)	2 (33.33%)
Other Rules		
Comma Rule	11 (58%)	8 (25%)
'OR' rule	10 (100%)	0
'AND' / 'OR' Rule	6 (86%)	1 (14.28%)
Pronoun Rule	9 (90%)	1 (10%)
Lexical Rule 1	9 (100%)	0
Lexical Rule 2	2 (50%)	2 (50%)
Two Conjunctions	2 (28.57%)	5 (71.42%)

3.3 Phrase-Phrase Rule:

Ifboth conjuncts are phrases, this rule is applied. This requires both syntactic and semantic analysis of the conjuncts.

A, B, C -- Z and A' B' C' --Z'
-if Z and Z' are nouns
Z and Z' are similar
or
-if A and A' are nouns / adjectives
A and A' are similar then
A B C -- Z OR A'B'C' --Z'

-if A and A' are dissimilar
Z and Z' are dissimilar
A B C --Z and A' B' C' --Z'
e.g., Christmas parties and drunken driving
Christmas parties AND Drunken driving

-if A and A' or
Z and Z' are identical (constitute the same words) then
A AND (B C --Z OR B' C' --Z')
e.g., Library Cooperation and library evaluation
Library (Cooperation OR Evaluation)

ZAND (A B C OR A' B' C')
e.g., Child daycare and elderly daycare
DaycareAND (Child OR Elderly)

3.4 Prepositional Phrase - Word Rule:

This rule handles sentences with the pattern "Prepositional phrase and word"; i.e., the first conjunct is a prepositional phrase and the second conjunct is a single word. This calls for the analysis of the semantic similarity of the last word in the prepositional phrase and the second conjunct. Based on the similarity or the dissimilarity, a Boolean 'OR' or 'AND' is introduced between the conjuncts. The preposition in the first conjunct is substituted by a Boolean 'AND'.

e.g., Aggressive behavior of handicapped children
and adolescents

AggressiveBehavior AND (Handicapped
Children OR Adolescents)

The same process is applied to sentences with the pattern "WORD and PREPOSITIONAL PHRASE".

3.5 Prepositional Phrase - Prepositional Phrase Rule:

This rule requires the analysis of semantic similarity between the last words in the prepositional phrases and also their syntactic analysis.

3.6 Phrase - Word Rule:

This applies to adjectival phrases and single words. Such patterns often require transformational rules to restructure the conjuncts, e.g., Depression treatment and diagnosis. This statement is transformed into 'depression treatment and depression diagnosis'. Then the similarity measure as in the phrase-phrase rule is applied to the conjuncts, which are indeed phrases after transformation.

-Depression AND (Treatment OR Diagnosis)

3.7 Word - Phrase Rule:

Phrase refers to adjectival phrases and they are run through a transformational process before the phrase-phrase rule is applied;

e.g., "Death and disability benefits" is transformed
to "Death benefits and disability benefits"
-BenefitsAND (Death OR Disability)

3.8 Lexical Rule 1:

This rule handles sentences consisting of the following words/phrases:

"effect of"
"impact of"
"relationship between"
"influence of"
"interaction between"
"interrelated"
"correlation between"

In all such cases the Boolean interpretation involves the use of the Boolean 'AND'.

e.g. Interaction between tropical agriculturists and
demographics
TropicalAgriculturistsAND Demographics

3.9 Lexical Rule 2:

Handles sentences with the following words / phrases:

"such as"
"like"
"specifically"
"for example"
"especially"

The Boolean 'OR' is used in such instances.

e.g. Antisocial personality such as psychopathic
personality and sociopathic personality.

Antisocial personality OR psychopathic
personality OR sociopathic personality

3.10 Pronoun Rule:

If a pronoun is present in a pattern then the conjuncts are combined with a Boolean 'AND' the pronoun is dropped.

e.g. Computers and their manufacture
ComputersAND Manufacture

3.11AND/OR, ORRule:

These conjunctions are interpreted as a Boolean 'OR'.
e.g. Compilers and/or computers
Compilers OR Computers

4. Boolean Interpreter

The "Boolean Interpreter" is an algorithm based on heuristics, which was developed for testing the rules. This was implemented on a PC in the 'C' programming language. The "Boolean Interpreter" accepts input either from the keyboard or from a file. Input is in the form of a series of sentences. Words in phrases are hyphenated to handle phrase scoping. Scoping, including syntactic and other methods, is beyond the scope of this study. Except commas, all punctuations are ignored. The input stringis analyzed and a list of tokens is built. In the process, articles 'a', 'an', and 'the', are eliminated. The program then applies the rules to the tokens. The tokens are analyzed left to right. Rules are applied in the order going from the most restrictive to the least restrictive ones. A pattern matching technique is used to determine the success or failure of the application of a rule to the input sentence. The conjuncts are transformed into Boolean operators. Processing is stopped when one of the following conditions is satisfied: - end of input tokens; -no more rules can be applied to transform the input. The program operates by means of a dialogue involving user input at different steps of the process, which it uses for further processing.

5. Evaluation of the Rules

For the purpose of evaluation, an expert searcher of online bibliographic databases was asked to formulate search expressions for the same set of 160 NLS and the expert's formulation was matched with the output of the algorithm. It resulted in an81% match rate. Table 3 gives the distribution of the matches.

Fifty-two or 32.5% of occurrences includes prepositional phrases, out of which 82.7% resulted in correct matches. The conjunction 'OR' was used in seventeen (10.62%) of the NLS. There was no occurrence of the conjunction 'but' in the NLS.

6. Observations

Definitions for the conjuncts in the NLS were largely found in LDOCE. Only in a marginal number of cases Webster's International Dictionary was used. There were some problems of identification of genus terms and their use. Different word forms of the same term occurred as genus terms in definitions; e.g., treatment, treats; common terms occurred as the genus terms and this posed some problems in tracing the hierarchy of definitions; levels of hierarchy tended to differ for the two conjuncts. The first conjunct might have required only one level of analysis of the definitions, while the second conjunct may have required more than one level. On an average, two levels of analysis were found adequate.

The rules, as of now cannot handle NLS with two conjuncts adequately; Phrase-Phrase rule needs to be worked on further.

7. Conclusion

The results are promising and suggest that this method when refined and developed further, could eventually be used in automatic Boolean interpretation of conjunctive phrases. Prior to this, research needs to be conducted towards adding, refining and developing the rules further, so as to accommodate variations in the NLS drawn from various disciplines, representing other phrase patterns, besides the ones tested in this study; the rules also need to be tested on larger samples of NLS.

Note:

* Expanded and revised version of a paper presented at KOTA'91, September '91 at Varna, Bulgaria.

References:

(1) Chodorow, M.S.; Byrd, R.J.; & Heidorn, G.E.(1985). Semantic hierarchies from a large on-line dictionary. Paper presented at 23rd Annual Meeting of the Assoc.Computat. Linguistics, Chicago, IL.

(2) Das-Gupta, P.: Boolean interpretation of conjunctions for document retrieval. J.Amer.Soc.Inform.Science 38(1987)No.4, p.245-254.

(3)Fox, E.A;Nutter, J.T.;Ahlswede, T.; Evens, M. Markowitz, J.: Building a large thesaurus for information retrieval. Paper 2nd Conf. on Applied Natural Language Processing, Assoc. Comput. Linguistics, Austin, TX, 1988.

(4) Longmans Dictionary of Contemporary English. Essex, England: Longman 1988.

(5) Markowitz, J.; Ahlswede, T.; Evans, M.: Semantically significant patterns in dictionary definitions. Paper 24th Ann. Meetg. Assoc. Comput. Linguistic. New York, NY, 1986.