Chapter 3
Concept Parsing Algorithms (CPA)

ABSTRACT

This chapter describe Concept Parsing Algorithms (CPA), a novel methodology of using text analysis tools for discovery of 'building blocks' of concepts, with semantic searches of the full text of potentially relevant documents in relevant knowledge domains, for lexical labels of concepts in controlled vocabularies. The meaning of lexical label of a super-ordinate concept C' in a sublanguage with controlled vocabulary is encoded in a set that contains three sets of building blocks: Ci (set of co-occurring sub-ordinate concepts); Rj (set of relations); and Lk (set of linguistic elements/descriptors).

INTRODUCTION

In the *Print Age* content was made accessible to library readers through curated paper card catalogues that included meta-tags encoded in Dewey Decimal Classification (DDC), that facilitate reader access to individual documents, namely, books, journals, reports, etc. DDC is a general knowledge organization tool (see: DDC 22 Summaries at oclc.org), where basic classes are organized by disciplines or fields of study. In the *Digital Age,* electronic catalogs provide the same meta-data electronically (see: Markey, 2015). Recent proliferation of digital books allows the inclusion in the functionality of electronic catalog a novel, game-changing feature of *Conceptual Curation* – semantic searches of the full text of potentially relevant documents in relevant knowledge domains

DOI: 10.4018/978-1-5225-2176-1.ch003

for names of concepts in controlled vocabularies. For a given 'super-ordinate' concept of interest, conceptual curation identifies collocated concepts, and reveal the conceptual context of a concept by identifying co-occurrences with other relevant concepts. Such co-occurrence may occur at the level of sentence, paragraph, page, or chapter. In a prescient view of semiotics of concepts, the linguist-semanticist de Beaugrande (1980) described conceptual meaning as embedded in a network of a knowledge domain as 'meaning of a concept is experienced by standing at its control center in a network and looking outward along all of its relational links in that knowledge space' (p. 68). In other words, de Beaugrande's description of 'conceptual meaning' foresaw it as facilitated by conceptual curation.

MAIN FOCUS OF THE CHAPTER

Issues, Controversies, Problems

In the following discussion we define as 'super-ordinate' a concept at the focus of interest, and denote it C'. Equation (3) is a set-theoretic definition of the super-ordinate concept C' in terms of its three building blocks, each containing a specific type of features/descriptors; this equation can be used as a generic format of *Concept Parsing Algorithms (CPA)* that guide the unpacking of C' into its component parts:

$$C' = \{[Ci], [Rj], [Lk]\} \quad (3)$$

The meaning of a lexical label of a super-ordinate concept C' in a sublanguage with controlled vocabulary is encoded in a set that contains three building blocks (Shafrir & Etkind, 2005; 2006); these are the sets:

[Ci] = set of co-occurring sub-ordinate concepts [C_1, C_2, …, Cm]
[Rj] = set of relations [R_1, R_2, …, Rn]
[Lk] = set of linguistic elements (descriptors) [L_1, L_2, …, Lp]

Sets [Ci], [Rj] and [Lk] in Equation (3) can be characterized by the following (non-exhaustive) list of descriptors:

Set [Ci] of Co-occurring Concepts: