

## 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:  $C_i$  (set of co-occurring sub-ordinate concepts);  $R_j$  (set of relations); and  $L_k$  (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 (Shafir & Etkind, 2005; 2006); these are the sets:

$[Ci]$  = set of co-occurring sub-ordinate concepts  $[C_1, C_2, \dots, C_m]$

$[Rj]$  = set of relations  $[R_1, R_2, \dots, R_n]$

$[Lk]$  = set of linguistic elements (descriptors)  $[L_1, L_2, \dots, L_p]$

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:

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/concept-parsing-algorithms-cpa/186506](http://www.igi-global.com/chapter/concept-parsing-algorithms-cpa/186506)

## Related Content

---

### Privacy Preserving Clustering for Distributed Homogeneous Gene Expression Data Sets

Xin Li (2012). *Innovations in Data Methodologies and Computational Algorithms for Medical Applications* (pp. 184-207).

[www.irma-international.org/chapter/privacy-preserving-clustering-distributed-homogeneous/65158](http://www.irma-international.org/chapter/privacy-preserving-clustering-distributed-homogeneous/65158)

### Development of Stress Management System for Prevention of a Mental Disease using Smart Finger Plethysmogram Measurement

Hidehiko Hayashi and Akinori Minazuki (2014). *International Journal of Computational Models and Algorithms in Medicine* (pp. 66-74).

[www.irma-international.org/article/development-of-stress-management-system-for-prevention-of-a-mental-disease-using-smart-finger-plethysmogram-measurement/148349](http://www.irma-international.org/article/development-of-stress-management-system-for-prevention-of-a-mental-disease-using-smart-finger-plethysmogram-measurement/148349)

### IR with and without GA: Study the Effectiveness of the Developed Fitness Function on the Two Suggested Approaches

Ammar Al-Dallal, Rasha S. Abdulwahab and Ramzi El-Haddadeh (2013). *International Journal of Applied Metaheuristic Computing* (pp. 1-20).

[www.irma-international.org/article/without-study-effectiveness-developed-fitness/77296](http://www.irma-international.org/article/without-study-effectiveness-developed-fitness/77296)

### The Use of Soft Computing for Optimization in Business, Economics, and Finance

Petr Dostál (2013). *Meta-Heuristics Optimization Algorithms in Engineering, Business, Economics, and Finance* (pp. 41-86).

[www.irma-international.org/chapter/use-soft-computing-optimization-business/69881](http://www.irma-international.org/chapter/use-soft-computing-optimization-business/69881)

### A New Mechanical Algorithm for Calculating the Amplitude Equation of the Reaction-Diffusion Systems

Houye Liu and Weiming Wang (2012). *International Journal of Computational Models and Algorithms in Medicine* (pp. 21-28).

[www.irma-international.org/article/a-new-mechanical-algorithm-for-calculating-the-amplitude-equation-of-the-reaction-diffusion-systems/101425](http://www.irma-international.org/article/a-new-mechanical-algorithm-for-calculating-the-amplitude-equation-of-the-reaction-diffusion-systems/101425)