§12. Automatic Enumeration of Scientific Text

Pichl, L. (ICU),
Kato, D.

Formalism of computer language is reviewed as a processing frame for science knowledge cataloguing, generation, and categorization.

Ever since the content depth of text in general can be measured structurally wise, starting from the elementary statistics of letters, counts of syllables, morphemes, words, relevant clause patterns, specific grammar categories, and onward, it is possible to establish the content equivalence patterns defined within such projection pattern spaces, an approach that is appealing in science especially due to its presumably high degree of notation and content hierarchy. Typing editors in the command based environments (such as latex) or those with graphical user interfaces (in markup tag representation of undertaken choices) can in principle provide a reasonable enumerating platform to facilitate such a type of a generic comparative measure. In addition, the emergence of semantic structures in text can be matched to generating grammars, adhering to the same shared representation template, whether the minimal size of Backus-Naur formal grammar, depth of the derived content parse tree in particular, or a suitable combination of both.

The production structure is expressed with < > to delimit arbitrary name labels, | for logical disjunction, ::= for unfolding assignment, and a pair of “ ” as text delimiters, in the standard form,

```
<syntax> ::= <rule> | <rule> <syntax>
<rule>::=<blanks>"<"<rule-name>">" <blanks>
   "::=" <blanks> <expression> <eol>
<blanks> ::= " " <blanks> | ""
<expression>::=<list> | <list> "|" <expression>
<eol> ::= <blanks> <ret> | <eol> <eol>
<list> ::= <term> | <term> <blanks> <list>
<term> ::= '"'<text>'"' | "<" <rule-name> ">"
```

where the single apostrophes denote a self-assertion mark of the frame connection junction, namely that the double apostrophes should be typed into the production rule carried inside. This annotation scheme instantiates not only on the markup tags of program input for complex simulation programs, but may apply to the output data as well, which relation in modular problems simply allows to imprint a causal relation onto the representation frame. Such results may then combine in segments, if an interface supporting a high degree of parallelism was adopted (e.g. molecular dynamics of plasma container material erosion). It should be noted here, however, that since notions (and problems) are usually established as rather well separated with respect to all the available means of knowledge, the applicability of structural depth assessment is theoretically interesting at the emergence of nonequilibrium phenomena, where the notion boundaries are stirred, such as the establishment of a turbulence, or the origin of a phase transition; runtime dataset tags may in addition be adjusted as necessary for the needs of simulation ensemble. Such a prospect perhaps directs towards the future processing architectures, where virtual machines assemble and merge on demand to meet the computing tasks of varying complexity.

Mathematical, notational and methodological imprints onto different fields and aspects in varying depth are nevertheless customary; a care should thus be taken to separate the depth of knowledge acquiring system from that of complexity generating system, in order to avoid effects of pervasive terminologically inefficient structures. Framed within the acquired resource structure, it is presumably possible to explore the irreducible constructive depth of knowledge segments for the sake of knowledge representation on demand.

In order to test this approach on a specific footing, we paraphrased the linguistic approach of multivariate knowledge partition through dictionaries and thesauri using purely stochastic content link generation, in a node link enumeration structure that is thereby asserted as the content meaning equivalent reduction [1]. This paraphrase of two or more stochastic languages in the abstracted way easily translates to the case of alternative simulation approaches, multiple experimental techniques, and so forth, subjected to comparative evaluation, and thus may provide a useful illustration to assess the expressive power of topological context of semantic graphs in computational science including plasma simulation.

The underlying components are classical in computer science, ranging from transitive closure, substance extraction, semantic extent, popularity ranking, similarity clustering, statistics of vertex degrees, content scaling behavior, or cycle detection, and thus assumed generally applicable and valid.

[1] L. Pichl, Semantic Network Closure Structures in Dual Translation of Stochastic Languages, S. Kikuchi, S. Sachdeva, and S. Bhalla (Eds.), Lecture Notes in Computer Science 5999 (2010) 210–224.