Generalized LL Parsing Generalization

Semyon Grigorev, Artyom Gorokhov

Saint Petersburg State University

7/9 Universitetskaya nab.

St. Petersburg, 199034 Russia

semen.grigorev@jetbrains.com, gorohov.art@gmail.com

Nowadays input data for parsing algorithms are not limited to be linear strings, and context-free grammars are used not only for programming languages specification. One of classical examples is a context-free path querying for graph data bases where input is a graph and path constraints are specified with grammar. There are also other generalizations of parsing, such as multiple input GLL parsing presented at Parsing@SLE-2016 by Elizabeth Scott and Adrian Johnstone, Abstract LR parsing [2] and other techniques for parsing of dynamically generated strings.

We have some experience in the mentioned areas [3, 5]. GLL-based context-free path querying algorithm [3] implemented by the authors is faster than solution which was presented at ISWC-2016 [6]. We have some ideas of graph parsing application. For example, context-free pattern search in metagenomical assemblies, which can be applied not only to regular input, but for context-free compressed input which is relevant for metagenomic assembly processing. All existing applications look like special cases of the Bar-Hillel [1] theorem for context-free and regular language intersection, and can be generalized, but today many of them are developed as stand alone solutions. Thus, the goal of our work is to create abstract framework for parsing based on geteralization of GLL parsing algorithm [4] proposed by Elizabeth Scott and Adrian Johnstone. We also want to investigate practical areas of application and to create solutions based on our framework to demonstrate its practical value.

References

[1] Bar-Hillel, Yehoshua, Micha Perles, and Eliahu Shamir. "On formal properties of simple phrase structure grammars." *Sprachtypologie und Universalienforschung* 14 (1961): 143-172.

[2] Doh, Kyung-Goo, Hyunha Kim, and David A. Schmidt. "Abstract LR-parsing.", *Formal Modeling: Actors, Open Systems, Biological Systems.*, Springer, 2011. 90–109.

[3] Grigorev, Semyon, and Anastasiya Ragozina. "Context-Free Path Querying with Structural Representation of Result." *arXiv preprint arXiv:1612.08872* (2016).

[4] Scott, Elizabeth, and Adrian Johnstone. "GLL parsing.", *Electronic Notes in Theoretical Computer Science*, 253.7 (2010): 177–189.

[5] Verbitskaia, Ekaterina, Semyon Grigorev, and Dmitry Avdyukhin. "Relaxed Parsing of Regular Approximations of String-Embedded Languages." *International Andrei Ershov Memorial Conference on Perspectives of System Informatics.* Springer International Publishing, 2015.

[6] Zhang, Xiaowang, et al. "Context-free path queries on RDF graphs." *International Semantic Web Conference.* Springer International Publishing, 2016. 632–648.