Graph parsing application for bio problems

Semyon Grigorev, Artem Gorokhov
Saint Petersburg State University
7/9 Universitetskaya nab.
St. Petersburg, 199034 Russia
semen.grigorev@jetbrains.com, gorohov.art@gmail.com

Biomedical databases contain vast amounts of rich data, much of which can be represented as a labeled graph. One of exampes is a raph where vertices correspond to entities and concepts labeled with their types such as gene, phenotype, and edges represent known relationships such as "codes for", "interacts with", etc. Paths between vertices may provide information about links were unknown before, forming the basis for new hypotheses.

Another example of graph structured data is metagenomic assemblies. Secondary structure can be described in terms of context-free grammar (Eddy et al), and grammar can be used for finding and classification. But for linera data. dispird the fact of tools existing, Graph structured data processing is still a challenge Context-free pattern search in metagenomical assemblies.

Analysis can be based on is a context-free path querying for graph data bases where input is a graph and path constraints are specified by a context-free grammar.

We have some experience in graph parsing [2, 4]. GLL-based context-free path querying algorithm [2] implemented by the authors is faster than solution which was presented at ISWC-2016 [5]. We have some ideas of graph parsing applications in bio data analysys.

References

[1] Sevon, Petteri, and Lauri Eronen. "Subgraph queries by context-free grammars." *Journal of Integrative Bioinformatics (JIB)* 5.2 (2008): 157-172.

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

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

[4] 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.

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